

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 05:50:54 ; Search time 535 Seconds  
(without alignments)  
9890.498 Million cell updates/sec

Title: US-10-617-962-2  
Perfect score: 1008  
Sequence: 1 atggttatacaattacacac.....ctcattataaataatataga 1008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*  
1: geneseqn1980s:\*  
2: geneseqn1908s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	100.0	1008	2	AAX07350 Photorhab
2	1008	100.0	1388	2	AAX07352 Photorhab
3	992	98.4	1008	10	ACF68075 Photorhab
4	992	98.4	110000	10	ACF67367_07 Continuation (8 of
5	992	98.4	110000	10	ACF65384_1 Continuation (2 of
6	663.8	65.9	1107	2	AAX07349 Xenorhabd
7	663.8	65.9	1205	2	AAX07351 Xenorhabd
8	628.2	62.3	1272	2	AAQ80839 Xenorhabd
9	52.6	5.2	3927	3	AAA70101 Plasmodiu
10	51	5.1	8056	8	ABZ10246 Haematopo
11	50.8	5.0	4590	1	AAW60472 Sequence
12	50.6	5.0	188971	12	ADL08108 Human gen
13	50.4	5.0	3738	3	AAW70178 Plasmodiu
14	49.8	4.9	11394	6	ABR28221 DNA trans
15	49.2	4.8	5940	3	AAA70105 Plasmodiu
16	48.8	4.8	1866	2	AAW99653 Nucleic a
17	48.2	4.8	2767	10	ADG42381 DNA encod
18	47.8	4.7	729	6	ABN91698 Staphyloc
19	47.6	4.7	1953	12	ADN19011 Staphyloc
20	47.6	4.7	16217	6	ABL32624 Human imm
21	47.4	4.7	717	8	ACF73065 Staphyloc

C 22	47.4	4.7	41708	3	AAW68247 Bacteriop
C 23	47.4	4.7	41708	4	AAW68106 Complete
C 24	47.4	4.7	158001	12	ADL17884 Human pho
C 25	47.2	4.7	700	4	AAW93026 Human inf
C 26	47	4.7	549	3	AAW93384 Cat flea
C 27	47	4.7	5340	5	AAW62829 P. falcip
C 28	47	4.7	11052	6	ABK39986 Human che
C 29	46.4	4.6	5409	3	AAW70151 Plasmodiu
C 30	46.4	4.6	8056	8	ABZ10100 Haematopo
C 31	46	4.6	32392	6	ABL56203 AmsPV gen
C 32	45.6	4.5	18283	6	ABL70502 Chemical1
C 33	45.6	4.5	18283	6	AAW61363 Human 3'
C 34	45.4	4.5	605	2	AAW31530 Human gen
C 35	45.4	4.5	9824	10	ADP42684 Plasmid v
C 36	45.2	4.5	4985	6	ABO75107 Anopheles
C 37	45.2	4.5	4985	10	ACF79720 Mosquito
C 38	45.2	4.5	13377	6	AAW46476 Tumour su
C 39	45.2	4.5	13377	6	ABL33463 Human imm
C 40	45	4.5	583	8	ADA72369 Rice gene
C 41	45	4.5	6103	6	ABL33691 Human imm
C 42	44.6	4.4	1488	4	AAW52007 Staphyloc
C 43	44.6	4.4	1488	8	ACF74231 Staphyloc
C 44	44.6	4.4	1491	4	AAW54922 Staphyloc
C 45	44.6	4.4	1491	8	ACA20304 Prokaryot

## ALIGNMENTS

RESULT 1  
ID AAX07350 standard; DNA; 1008 BP.

AC AAX07350;

DT 21-MAY-1999 (first entry)

DE Photorhabdus luminescens insecticidal toxin gene P1V16tox1.

XX Toxin; P1V16tox1; biological control; insecticide; ds.

OS Photorhabdus luminescens.

PN WO9903328-A1.

XX 28-JAN-1999.

PF 17-JUL-1998; 98WO-AU000562.

XX 17-JUL-1997; 97AU-0008088.

PR 17-JUL-1997; 97AU-0008088.

PA (CSTR ) COMMONWEALTH SCI & IND RES ORG.

PI East PD;

XX WPI; 1999-131733/11.

DR P-PSDB; AAW97811.

PT New insecticidal toxin genes - extracted from Xenorhabdus nematophilus

XX A24 and Photorhabdus luminescens.

XX Claim 1; Page 27-28; 48pp; English.

This is the coding region of the insecticidally-active toxin gene, termed P1V16tox1, of Photorhabdus luminescens strain V16/1. It codes for a 335-amino acid P1V16tox1 protein (see AAW97811). The gene was isolated from a genomic DNA cosmid library using the Xenorhabdus nematophilus A24 tox24 gene (see AAX07349) as probe. A 1.39 kb HindIII/SmaI fragment (see AAX07352) of the isolated gene included the 1008 bp coding region. The invention relates to the identification and isolation of polynucleotide molecules encoding a new class of protein insecticidal toxins which are produced by bacteria of the genera Xenorhabdus and Photorhabdus. The polynucleotide molecules may be incorporated e.g. into insect-specific

CC viruses (e.g. entomopox and nuclear polyhedrosis viruses), bacteria (e.g. CC Gracilicutes, Firmicutes, Tenericutes and Mendosicutes), protozoa, yeast CC and plants (particularly cereals such as wheat and barley, vegetables CC such as tomato and potato, fruit trees such as citrus and apple, tobacco CC and cotton) for control of pest insects

XX Sequence 1008 BP; 376 A; 146 C; 192 G; 294 T; 0 U; 0 Other;

Query Match 100.0%; Score 1008; DB 2; Length 1008;  
Best Local Similarity 100.0%; Pred. No. 2e-210;  
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGGTTATACATTAACACCTGATGATAGAGTATATCCACCCGTTGAAAAACAATA 60
DB 1 ATGGTTATACATTAACACCTGATGATAGAGTATATCCACCCGTTGAAAAACAATA 60
QY 61 GGAGAGATATAGTACGTTACTTAACTTAAGCAAAAGAGAGGTCATCAGACATCA 120
DB 61 GGAGAGATATAGTACGTTACTTAACTTAAGCAAAAGAGAGGTCATCAGACATCA 120
QY 121 TATGGAATGATATCGAGCTAAGAAATAATATTAGCTTACGCTTGGCTGTAAGTGT 180
DB 121 TATGGAATGATATCGAGCTAAGAAATAATATTAGCTTACGCTTGGCTGTAAGTGT 180
QY 181 ATTCAATATGATCTAACTTCTGATGACATATATATAGAAATPAAAGACTGCTAGAGA 240
DB 181 ATTCAATATGATCTAACTTCTGATGACATATATATAGAAATPAAAGACTGCTAGAGA 240
QY 241 ATTTATCAAGATATATGCTTATCTTATCATCTGACATTAATGAGGAAATGGTATCA 300
DB 241 ATTTATCAAGATATATGCTTATCTTATCATCTGACATTAATGAGGAAATGGTATCA 300
QY 301 ATTTCTAAAGATATGCAATGCTTTTATTAAGATGAACTGATTTGMAAGTCAATAT 360
DB 301 ATTTCTAAAGATATGCAATGCTTTTATTAAGATGAACTGATTTGMAAGTCAATAT 360
QY 361 CCTCAAAACATTGGAAATGCTTCTGAGCTTGAATAAACAATGATGCTTATTCAGAT 420
DB 361 CCTCAAAACATTGGAAATGCTTCTGAGCTTGAATAAACAATGATGCTTATTCAGAT 420
QY 421 GACGATTAATTTTGGCACTATATTTTCTGACAGAAATTCACATGAGAGAAAT 480
DB 421 GACGATTAATTTTGGCACTATATTTTCTGACAGAAATTCACATGAGAGAAAT 480
QY 481 CAACATCAAAATGCCGCAAGATTTTAAATTAATGATTTCTTATTTACTTATCTGCT 540
DB 481 CAACATCAAAATGCCGCAAGATTTTAAATTAATGATTTCTTATTTACTTATCTGCT 540
QY 541 GTAACCTTCACTGGGAAAGAGATTTTCAAAAACCTTTTACATGATTAAGGCTTAA 600
DB 541 GTAACCTTCACTGGGAAAGAGATTTTCAAAAACCTTTTACATGATTAAGGCTTAA 600
QY 601 TCATTAGGAATTTATTTAGAGAAATAAATCTTTCAACCTTTCTTTCGACACCGGAG 660
DB 601 TCATTAGGAATTTATTTAGAGAAATAAATCTTTCAACCTTTCTTTCGACACCGGAG 660
QY 661 AGATTACCTGATGAGAGATAGATTATTTGCTGACCAACAGAAAGCGCTTAATGAGA 720
DB 661 AGATTACCTGATGAGAGATAGATTATTTGCTGACCAACAGAAAGCGCTTAATGAGA 720
QY 721 GTGAGTTTAAAGAACTTAAATAAACAATCTAGAGATGATTTCTTATATGAGAGG 780
DB 721 GTGAGTTTAAAGAACTTAAATAAACAATCTAGAGATGATTTCTTATATGAGAGG 780
QY 781 GCTGAAAACAAAGATATGTCATTTTAAAGAGGTCAAAAGGTTACGCTCCACAG 840
DB 781 GCTGAAAACAAAGATATGTCATTTTAAAGAGGTCAAAAGGTTACGCTCCACAG 840
QY 841 ACAGACGAGAAAGATTTGGTACAGCAGTGCAGATTAATCTGAAAAATTTGCCAATAT 900
DB 841 ACAGACGAGAAAGATTTGGTACAGCAGTGCAGATTAATCTGAAAAATTTGCCAATAT 900
QY 901 TTATATAGTGTAGGCTTAAGCCAAAAAGACAGGGTAACTTTATCTCAAAATGATCTGAC 960

```

```

DB 901 TTATATAGTGTAGGCTTAAGCCAAAAAGACAGGGTAACTTTACTCAAAAAGATATCTGAC 960
QY 961 AATTCATGACGTTTCATAGTGTGGAACCTCATTTAAAAATATATATGA 1008
DB 961 AATTCATGACGTTTCATAGTGTGGAACCTCATTTAAAAATATATATGA 1008

```

RESULT 2

AXX07352  
ID AAX07352 standard; DNA; 1388 BP.

XX AAX07352;

XX 21-MAY-1999 (first entry)

XX Photorhabdus luminescens insecticidal toxin gene pVI6tox1.

XX Toxin; pVI6tox1; biological control; insecticide; de.

XX Photorhabdus luminescens.

XX Key Location/Qualifiers

FT CDS 172..1179

FT /\*tag= a

FT /note= "the coding region is specifically claimed in

PN MO9903328-A1.

PD 28-JAN-1999.

XX 17-JUL-1998; 98WO-AU000562.

XX 17-JUL-1997; 97AU-00008088.

XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.

XX East PD;

DR WPI, 1999-131733/11.

DR P-PSDB; AAX97811.

XX New insecticidal toxin genes - extracted from Xenorhabdus nematophilus

XX A24 and Photorhabdus luminescens.

XX Example 1; Page 32; 48pp; English.

CC This is the nucleotide sequence of the insecticidally-active toxin gene, CC termed pVI6tox1, of Photorhabdus luminescens strain V16/1. It includes a CC coding region, which is claimed (see AAX07350), that codes for a 335- CC amino acid pVI6tox1 protein (see AAX97811). The gene was isolated from a CC genomic DNA cosmid library using the Xenorhabdus nematophilus A24 toxB4 CC gene (see AAX07349) as probe. The invention relates to the identification CC and isolation of polynucleotide molecules encoding a new class of protein CC insecticidal toxins which are produced by bacteria of the genera CC Xenorhabdus and Photorhabdus. The polynucleotide molecules may be CC incorporated e.g. into insect-specific viruses (e.g. entomopox and CC nuclear polyhedrosis viruses), bacteria (e.g. Gracilicutes, Firmicutes, CC Tenericutes and Mendosicutes), protozoa, yeast and plants (particularly CC cereals such as wheat and barley, vegetables such as tomato and potato, CC fruit trees such as citrus and apple, tobacco and cotton) for control of CC pest insects

XX Sequence 1388 BP; 501 A; 213 C; 261 G; 413 T; 0 U; 0 Other;

Query Match 100.0%; Score 1008; DB 2; Length 1388;  
Best Local Similarity 100.0%; Pred. No. 2.1e-210;  
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGGTTATACATTAACACCTGATGATAGAGTATATCCACCCGTTGAAAAACAATA 60
DB 172 ATGGTTATACATTAACACCTGATGATAGAGTATATCCACCCGTTGAAAAACAATA 231

```

```

QY 61 GCAGAGATATAGTACGTATCTAACTTTAAGCAACAGATGAGGCTATACGCAATCA 120
DB 232 GCAGAGATATAGTACGTATCTAACTTTAAGCAACAGATGAGGCTATACGCAATCA 291
QY 121 TATGGAATTGATATCGAGCTAGAGAAATTAATATTAGCTTCGCTTGGCTGAAGTGT 180
DB 292 TATGGAATTGATATCGAGCTAGAGAAATTAATATTAGCTTCGCTTGGCTGAAGTGT 351
QY 181 ATTCAATATGATATCTAACTTCCTGATGATCTATTATGAATTAAGAGACTGCTGAGGA 240
DB 352 ATTCAATATGATATCTAACTTCCTGATGATCTATTATGAATTAAGAGACTGCTGAGGA 411
QY 241 ATTATCAAGAAATATATGCTATATCTTCATCTGCACTATTAAGGAAATAGTGAATCA 300
DB 412 ATTATCAAGAAATATATGCTATATCTTCATCTGCACTATTAAGGAAATAGTGAATCA 471
QY 301 ATTCTTAAGATATATGCAATATGCTTTTATTAAGATGAATGCAATTTTGAAGTCAATAT 360
DB 472 ATTCTTAAGATATATGCAATATGCTTTTATTAAGATGAATGCAATTTTGAAGTCAATAT 531
QY 361 CCTCAAAACATTTGGAATGTTCTGAGCTTGAAATTAACCAATGAGTCTTATTCAGAT 420
DB 532 CCTCAAAACATTTGGAATGTTCTGAGCTTGAAATTAACCAATGAGTCTTATTCAGAT 591
QY 421 GACGATAAATATATGCACTATATTTTCTCTGACAGGAAATCCACTGGAGAAAT 480
DB 592 GACGATAAATATATGCACTATATTTTCTCTGACAGGAAATCCACTGGAGAAAT 651
QY 481 CAACATCAATATGCGCAAGATTTTATTAATTAATTTGATTTTACCTTATCTGCT 540
DB 652 CAACATCAATATGCGCAAGATTTTATTAATTAATTTGATTTTACCTTATCTGCT 711
QY 541 GTAACCTCACTGGAGAGAGATTTTTCAAAAAATTTTCAATGATGATGAGGCTAAA 600
DB 712 GTAACCTCACTGGAGAGAGATTTTTCAAAAAATTTTCAATGATGATGAGGCTAAA 771
QY 601 TCATTTAGAAATTAATATGAGAGAAATTAATTTTCAATGATGATGAGGCTAAA 660
DB 772 TCATTTAGAAATTAATATGAGAGAAATTAATTTTCAATGATGATGAGGCTAAA 831
QY 661 AGATTACCTGATGAGAGATATGATTTTGGCTGAGCAACAGAGCCCTTAATGAGGA 720
DB 832 AGATTACCTGATGAGAGATATGATTTTGGCTGAGCAACAGAGCCCTTAATGAGGA 891
QY 721 GTGAGTTTAAAGAACTTAAATTAACAATCTAGGAATGATTTTCTTAATATGAGAGGG 780
DB 892 GTGAGTTTAAAGAACTTAAATTAACAATCTAGGAATGATTTTCTTAATATGAGAGGG 951
QY 781 GCTGCAAAACAAATATATGATTTTCAATTTAAAGAGGTRCAAAAGGTRACGCTCCAGG 840
DB 952 GCTGCAAAACAAATATATGATTTTCAATTTAAAGAGGTRCAAAAGGTRACGCTCCAGG 1011
QY 841 ACAGAGAGGAAAGATTTGTTAGTACAGCCAGTGGCACTTAACCTGAGAAATATGCGCAATAT 900
DB 1012 ACAGAGAGGAAAGATTTGTTAGTACAGCCAGTGGCACTTAACCTGAGAAATATGCGCAATAT 1071
QY 901 TTATATATGTTGAGGCTTAAGCCAAAGAGAGGTTACCTTTACTCAAAATGATATCTGAC 960
DB 1072 TTATATATGTTGAGGCTTAAGCCAAAGAGAGGTTACCTTTACTCAAAATGATATCTGAC 1131
QY 961 AATATCAATGAGGCTTCAATATGTTGGAATCTCATTTAAATTAATATATGGA 1008
DB 1132 AATATCAATGAGGCTTCAATATGTTGGAATCTCATTTAAATTAATATATGGA 1179

```

RESULT 3  
 ACF68075 standard; DNA; 1008 BP.  
 AC ACF68075;  
 XX  
 DT 20-NOV-2003 (first entry)

```

XX XX Photorhabdus luminescens nucleotide sequence #6542.
DE Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX OS Photorhabdus luminescens.
XX WO200294867-A2.
XX 28-NOV-2002.
XX 07-FEB-2002; 2002WO-1B003040.
XX 07-FEB-2001; 2001FR-00001659.
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A,
XX Buchrieser C;
XX WPI; 2003-148459/14.
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX Claim 2; SEQ ID NO 6542; 1205bp; French.
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens genes
XX
XX Sequence 1008 BP; 377 A; 147 C; 191 G; 293 T; 0 U; 0 Other;
XX
XX Query Match 98.4%; Score 992; DB 10; Length 1008;
XX Best Local Similarity 99.0%; Pred. No. 6.2e-207; Indels 0; Gaps 0;
XX Matches 998; Conservative 0; Mismatches 10;
XX
XX QY 1 ATGGTTATACATTAATTAACCTGATGATGATGATATTCACCCGTGAAAGCAATA 60
DB 1 ATGGTTATACATTAATTAACCTGATGATGATGATGATATTCACCCGTGAAAGCAATA 60
QY 61 GCAGAGATATAGTACGTATCTAACTTTAAGCAACAGATGAGGCTATACGCAATCA 120
DB 61 GCAGAGATATAGTACGTATCTAACTTTAAGCAACAGATGAGGCTATACGCAATCA 120
QY 121 TATGGAATTGATATCGAGCTAGAGAAATTAATATTAGCTTCGCTTGGCTGAAGTGT 180
DB 121 TATGGAATTGATATCGAGCTAGAGAAATTAATATTAGCTTCGCTTGGCTGAAGTGT 180
QY 181 ATTCAATATGATATCTAACTTCCTGATGATCTATTATGAATTAAGAGACTGCTGAGGA 240

```

Db 181 ATTCAATATGATCTAAACTTCTGATGACTATTATAGAAATTAAGACCTGCTGAGAGA 240  
Qy 241 ATTATCAAGAAATATATGTCATATCTTTCATCTGCACTATTAGTGAAATGATGATCA 300  
Db 241 ATTATCAAGAAATATATGTCATATCTTTCATCTGCACTATTAGTGAAATGATGATCA 300  
Qy 301 ATTTCTAAAGATATGCAAAATGTTTATTAAGATGAATGATGATTTTGAAGTCATAT 360  
Db 301 ATTTCTAAAGATATGCAAAATGTTTATTAAGATGAATGATGATTTTGAAGTCATAT 360  
Qy 361 CCTGAAAACATTTGGAAATGTCCTGAGCTTGAAATTAACATGAGTCTTATTCAGAT 420  
Db 361 CCTGAAAACATTTGGAAATGTCCTGAGCTTGAAATTAACATGAGTCTTATTCAGAT 420  
Qy 421 GACGATTAATTAATGACACTATATTTTCTCTGACAGAAATTCACATGAGGAAAT 480  
Db 421 GACGATTAATTAATGACACTATATTTTCTCTGACAGAAATTCACATGAGGAAAT 480  
Qy 481 CACGATCAAAATGCGCGAAGATTTTAAATTAATGATTTCTATTTACCTTATCTGCT 540  
Db 481 CACGATCAAAATGCGCGAAGATTTTAAATTAATGATTTCTTACTTATCTTATCTGCT 540  
Qy 541 GTTACTTCACTGGAGAGAGATTTTTCAAAAAACTTTTACAAATGATTAAGGCTTAA 600  
Db 541 GTTACTTCACTGGAGAGAGATTTTTCAAAAAACTTTTACAAATGATTAAGGCTTAA 600  
Qy 601 TCATTAGAGAAATTAATTAATGAGAAAAAACTTTCTTAAACCTTTCTTTCGACACCGCAG 660  
Db 601 TCATTAGAGAAATTAATTAATGAGAAAAAACTTTCTTAAACCTTTCTTTCGACACCGCAG 660  
Qy 661 AGATTACCTGATGAGCAATATGATTTTGGCTGACCAAGAAAGGCTTAAATGAGAGA 720  
Db 661 AGATTACCTGATGAGCAATATGATTTTGGCTGACCAAGAAAGGCTTAAATGAGAGA 720  
Qy 721 GTGAGTTTAAAGAACTTAAATTAACAATCTAGAGATGATTTTCTATATGAGAGG 780  
Db 721 GTGAGTTTAAAGAACTTAAATTAACAATCTAGAGATGATTTTCTATATGAGAGG 780  
Qy 781 GCTGCAAAAAGAAAGTATGATTCATTATTAAGAGATCAAAAGGATTAACGCTCCACAG 840  
Db 781 GCTGCAAAAAGAAAGTATGATTCATTATTAAGAGATCAAAAGGATTAACGCTCCACAG 840  
Qy 841 ACAGAGAGGAAAGATTTGATTCAGCCAGTGGCAGTAACTTGGAAAAATTCGCCGATAT 900  
Db 841 ACAGAGAGGAAAGATTTGATTCAGCCAGTGGCAGTAACTTGGAAAAATTCGCCGATAT 900  
Qy 901 TTATATAGTGTAGGCTAAGCCAAAAAGCAGGGTAACTTACTCAAAATGATATCTGAC 960  
Db 901 TTATATAGTGTAGGCTAAGCCAAAAAGCAGGGTAACTTACTCAAAATGATATCTGAC 960  
Qy 961 AATACAAATGACGGTTCATAGTGTGGAACCTCATTAATAAATATATATGA 1008  
Db 961 AATACAAATGACGGTTCATAGTGTGGAACCTCATTAATAAATATATATGA 1008

RESULT 4  
ACF67367\_07/c  
Continuation (8 of 57) of ACF67367 from base 700001 (Photobhabdus luminescens nucleotide  
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367  
Fragment Name Begin End  
WP ACF67367\_00 1 110000  
WP ACF67367\_01 100001 210000  
WP ACF67367\_02 200001 310000  
WP ACF67367\_03 300001 410000  
WP ACF67367\_04 400001 510000  
WP ACF67367\_05 500001 610000  
WP ACF67367\_06 600001 710000  
WP ACF67367\_07 700001 810000  
WP ACF67367\_08 800001 910000  
WP ACF67367\_09 900001 1010000  
WP ACF67367\_10 1000001 1110000  
WP ACF67367\_11 1100001 1210000  
WP ACF67367\_12 1200001 1310000

WP ACF67367\_13 1300001 1410000  
WP ACF67367\_14 1400001 1510000  
WP ACF67367\_15 1500001 1610000  
WP ACF67367\_16 1600001 1710000  
WP ACF67367\_17 1700001 1810000  
WP ACF67367\_18 1800001 1910000  
WP ACF67367\_19 1900001 2010000  
WP ACF67367\_20 2000001 2110000  
WP ACF67367\_21 2100001 2210000  
WP ACF67367\_22 2200001 2310000  
WP ACF67367\_23 2300001 2410000  
WP ACF67367\_24 2400001 2510000  
WP ACF67367\_25 2500001 2610000  
WP ACF67367\_26 2600001 2710000  
WP ACF67367\_27 2700001 2810000  
WP ACF67367\_28 2800001 2910000  
WP ACF67367\_29 2900001 3010000  
WP ACF67367\_30 3000001 3110000  
WP ACF67367\_31 3100001 3210000  
WP ACF67367\_32 3200001 3310000  
WP ACF67367\_33 3300001 3410000  
WP ACF67367\_34 3400001 3510000  
WP ACF67367\_35 3500001 3610000  
WP ACF67367\_36 3600001 3710000  
WP ACF67367\_37 3700001 3810000  
WP ACF67367\_38 3800001 3910000  
WP ACF67367\_39 3900001 4010000  
WP ACF67367\_40 4000001 4110000  
WP ACF67367\_41 4100001 4210000  
WP ACF67367\_42 4200001 4310000  
WP ACF67367\_43 4300001 4410000  
WP ACF67367\_44 4400001 4510000  
WP ACF67367\_45 4500001 4610000  
WP ACF67367\_46 4600001 4710000  
WP ACF67367\_47 4700001 4810000  
WP ACF67367\_48 4800001 4910000  
WP ACF67367\_49 4900001 5010000  
WP ACF67367\_50 5000001 5110000  
WP ACF67367\_51 5100001 5210000  
WP ACF67367\_52 5200001 5310000  
WP ACF67367\_53 5300001 5410000  
WP ACF67367\_54 5400001 5510000  
WP ACF67367\_55 5500001 5610000  
WP ACF67367\_56 5600001 5648894

Query Match 98.4%; Score 992; DB 10; Length 110000;  
Best Local Similarity 99.0%; Pred. No. 1.5e-206;  
Matches 998; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 ATGTTTATCAATTAATTAACCTGATGATGAAGTGAATATCCACCGTTGAAAAAGCAATTA 60  
Db 66005 ATGTTTATCAATTAATTAACCTGATGATGAAGTGAATATCCACCGTTGAAAAAGCAATTA 65946  
Qy 61 GCAGAGATATATAGTACGATATCTTAACTTTAAGCAAAAGATGAGGTCATACAGCATCA 120  
Db 65945 GCAGAGATATATAGTACGATATCTTAACTTTAAGCAAAAGATGAGGTCATACAGCATCA 65886  
Qy 121 TATGAAATTAATATCGAGTAAAGAAATTAATATTAAGCTTAACGCTTGGCTGTAAGTGT 180  
Db 65885 TATGAAATTAATATCGAGTAAAGAAATTAATATTAAGCTTAACGCTTGGCTGTAAGTGT 65826  
Qy 181 ATTCATTAATGATCTTAAACCTTCTGATGACTATTAATAAATAAAGAGACTGCTGAGAGA 240  
Db 65825 ATTCATTAATGATCTTAAACCTTCTGATGACTATTAATAAATAAAGAGACTGCTGAGAGA 65766  
Qy 241 ATTTATCAAGAAATATATGTCATATCTTTCATCTGCACTATTAGTGAAATGATGATCA 300  
Db 65765 ATTTATCAAGAAATATATGTCATATCTTTCATCTGCACTATTAGTGAAATGATGATCA 65706  
Qy 301 ATTTCTAAAGATATGCAAAATGTTTATTAAGATGAATGATGATTTTGAAGTCATAT 360  
Db 65705 ATTTCTAAAGATATGCAAAATGTTTATTAAGATGAATGATGATTTTGAAGTCATAT 65646



QY 361 CCTCAAAACATTGGATGTTCTTGAGCTGTGAATAAACAATGAGCTTATTCAGAT 420  
Db 65645 CCTCAAAACATTGGATGTTCTTGAGCTGTGAATAAACAATGAGCTTATTCAGAT 65586  
QY 421 GACGATTAATTAATGACACTATATTTTCTCTGTACAGAAATCCACTGGAGAAAT 480  
Db 65585 GACGATTAATTAATGACACTATATTTTCTCTGTACAGAAATCCACTGGAGAAAT 65526  
QY 481 CAACAATCAATGCGGCAAGATTTTAAATTAATGATTTCTTAACTGATCTGCT 540  
Db 65525 CAACAATCAATGCGGCAAGATTTTAAATTAATGATTTCTTAACTGATCTGCT 65466  
QY 541 GTAACCTCACTGGAGAGAGATTTTCAAAAACTTTTCAATGATGAGGCTTAA 600  
Db 65465 GTAACCTCACTGGAGAGAGATTTTCAAAAACTTTTCAATGATGAGGCTTAA 65406  
QY 601 TCATTAGAAATTAATTAATGAGAAAAAACTTTCTTAACTTTTCTGACCAACGCGAG 660  
Db 65405 TCATTAGAAATTAATTAATGAGAAAAAACTTTCTTAACTTTTCTGACCAACGCGAG 65346  
QY 661 AGATTACCTGATGCGAGAAATGTTATTTGCTGACCAACGAGCGGCTTAAATGAGA 720  
Db 65345 AGATTACCTGATGCGAGAAATGTTATTTGCTGACCAACGAGCGGCTTAAATGAGA 65286  
QY 721 GTGAGTTTAAAGAACTTAAATAACAATCTAGGAATGATTTCTTAAATGAGAGG 780  
Db 65285 GTGAGTTTAAAGAACTTAAATAACAATCTAGGAATGATTTCTTAAATGAGAGG 65226  
QY 781 GCTGCAAAAACAAAGATATAGTTCAATTATTAAGAAGTACAAAGGGTACGCTCCACAG 840  
Db 65225 GCTGCAAAAACAAAGATATAGTTCAATTATTAAGAAGTACAAAGGGTACGCTCCACAG 65166  
QY 841 ACAGAGCGGAAAGATATGCTGACAGCCAGCTGACATTAACCTGGAATAATTCGCCAATAT 900  
Db 65165 ACAGAGCGGAAAGATATGCTGACAGCCAGCTGACATTAACCTGGAATAATTCGCCAATAT 65106  
QY 901 TTATATAGTGTGAGGCTTAAGCCAAAAAGACAGGTAACCTTACTCAAAATGATACTGAC 960  
Db 65105 TTATATAGTGTGAGGCTTAAGCCAAAAAGACAGGTAACCTTACTCAAAATGATACTGAC 65046  
QY 961 AATACAATGACGGTTCATAGTGTGAACTCATTTATTAATAATATATGA 1008  
Db 65045 AATACAATGACGGTTCATAGTGTGAACTCATTTATTAATAATATATGA 64998

RESULT 5  
ACF65384\_1/c  
Continuation (2 of 6) of ACF65384 from base 100001 (Photobhabdus luminescens nucleotide  
WP Sequence split into 6 fragments LOCUS ACF65384 Accession ACF65384  
Fragment Name Begin End  
WP ACF65384\_0 1 110000  
WP ACF65384\_1 100001 210000  
WP ACF65384\_2 200001 310000  
WP ACF65384\_3 300001 410000  
WP ACF65384\_4 400001 510000  
WP ACF65384\_5 500001 530312

Query Match 98.4%; Score 992; DB 10; Length 110000;  
Best Local Similarity 99.0%; Pred. No. 1.5e-206;  
Matches 998; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1 ATGGTTATACATTAACACCTGATAGTATGAGTATTCACCGGTTGAAAGCAATA 60  
Db 66220 ATGGTTATACATTAACACCTGATAGTATGAGTATTCACCGGTTGAAAGCAATA 66161  
QY 61 GCAGAGATATATGATGCTATTAACCTTAAAGCAACGATGAGGTCATACACATCA 120  
Db 66160 GCAGAGATATATGATGCTATTAACCTTAAAGCAACGATGAGGTCATACACATCA 66101  
QY 121 TATGAAATTTGATATGAGCTTAAGAAATAATATTAAGCTTATGCTTTGCTGTAGTGT 180  
Db 66100 TATGAAATTTGATATGAGCTTAAGAAATAATATTAAGCTTATGCTTTGCTGTAGTGT 66041

QY 181 ATTCAATATGATCTAACTTCTGATGACTATTAATAAGATTAAGAGACTGCTGAGAGA 240  
Db 66040 ATTCAATATGATCTAACTTCTGATGACTATTAATAAGATTAAGAGACTGCTGAGAGA 65981  
QY 241 ATTATCAAGATATATGCTTAATCTTCAATCTGACATTAATGAGTGAATAATGCTGATCA 300  
Db 65980 ATTATCAAGATATATGCTTAATCTTCAATCTGACATTAATGAGTGAATAATGCTGATCA 65921  
QY 301 ATTCTTAAGATATGCGCAAGATGTTTTTAAAGATGAATGAACTGATTTGAAAGTCAATAT 360  
Db 65920 ATTCTTAAGATATGCGCAAGATGTTTTTAAAGATGAATGAACTGATTTGAAAGTCAATAT 65861  
QY 361 CCTCAAAACATTGGAAATGTTCTTGAGCTGTGAATAAACAATGAGCTTATTCAGAT 420  
Db 65860 CCTCAAAACATTGGAAATGTTCTTGAGCTGTGAATAAACAATGAGCTTATTCAGAT 65801  
QY 421 GACGATTAATTAATGACACTATATTTTCTCTGTACAGAAATCCACTGGAGAAAT 480  
Db 65800 GACGATTAATTAATGACACTATATTTTCTCTGTACAGAAATCCACTGGAGAAAT 65741  
QY 481 CAACAATCAATGCGGCAAGATTTTAAATTAATGATTTCTTAACTTAACTGATCTGCT 540  
Db 65740 CAACAATCAATGCGGCAAGATTTTAAATTAATGATTTCTTAACTTAACTGATCTGCT 65681  
QY 541 GTAACCTCACTGGAGAGAGATTTTCAAAAACTTTTCAATGATGAGGCTTAA 600  
Db 65680 GTAACCTCACTGGAGAGAGATTTTCAAAAACTTTTCAATGATGAGGCTTAA 65621  
QY 601 TCATTAGAAATTAATTAATGAGAAAAAACTTTCTTAACTTTTCTTCTGACCAACGCGAG 660  
Db 65620 TCATTAGAAATTAATTAATGAGAAAAAACTTTCTTAACTTTTCTTCTGACCAACGCGAG 65561  
QY 661 AGATTACCTGATGCGAGAAATGTTATTTGCTGACCAACGAGCGGCTTAAATGAGA 720  
Db 65560 AGATTACCTGATGCGAGAAATGTTATTTGCTGACCAACGAGCGGCTTAAATGAGA 65501  
QY 721 GTGAGTTTAAAGAACTTAAATAACAATCTAGGAATGATTTCTTAAATGAGAGG 780  
Db 65500 GTGAGTTTAAAGAACTTAAATAACAATCTAGGAATGATTTCTTAAATGAGAGG 65441  
QY 781 GCTGCAAAAACAAAGATATAGTTCAATTATTAAGAAGTACAAAGGGTACGCTCCACAG 840  
Db 65440 GCTGCAAAAACAAAGATATAGTTCAATTATTAAGAAGTACAAAGGGTACGCTCCACAG 65381  
QY 841 ACAGAGCGGAAAGATATGCTGACAGCCAGCTGACATTAACCTGGAATAATTCGCCAATAT 900  
Db 65380 ACAGAGCGGAAAGATATGCTGACAGCCAGCTGACATTAACCTGGAATAATTCGCCAATAT 65321  
QY 901 TTATATAGTGTGAGGCTTAAGCCAAAAAGACAGGTAACCTTACTCAAAATGATACTGAC 960  
Db 65320 TTATATAGTGTGAGGCTTAAGCCAAAAAGACAGGTAACCTTACTCAAAATGATACTGAC 65261  
QY 961 AATACAATGACGGTTCATAGTGTGAACTCATTTATTAATAATATATGA 1008  
Db 65260 AATACAATGACGGTTCATAGTGTGAACTCATTTATTAATAATATATGA 65213

RESULT 6  
AAK07349  
ID AAK07349 standard; DNA; 1107 BP.  
XX  
XX AAK07349;  
XX AC  
XX 17-OCT-2003 (revised)  
XX DT 21-MAY-1999 (first entry)  
XX DE  
XX Xenorhabdus nematophilus insecticidal toxin gene toxB4.  
XX Toxin; toxB4; biological control; insecticide; ds.  
XX Xenorhabdus nematophilus.  
XX OS  
XX PN W09903328-A1.

```

XX 28-JAN-1999.
PD 17-JUL-1998; 98WO-AU000562.
XX PF 17-JUL-1997; 97AU-00008088.
XX PR 17-JUL-1997; 97AU-00008088.
XX PA (CSTR ) COMMONWEALTH SCI & IND RES ORG.
XX East PD;
XX MPI; 1999-131733/11.
XX DR P-PSDB; AAM97810.
XX PT New insecticidal toxin genes - extracted from Xenorhabdus nematophilus
XX A24 and Photorhabdus luminescens.
XX PS Claim 1; Page 25-26; 48pp; English.
XX
CC This is the coding region of the insecticidally-active toxin gene, termed
CC toxb4, of Xenorhabdus nematophilus strain A24. It codes for a 368-amino
CC acid toxb4 protein (see AAM97810). The gene was isolated by functional
CC screening of a bacterial DNA library using Galleria mellonella fourth
CC instar larvae. The isolated toxb4 clone contained an insert of 1205 bp
CC (see AAX07351) that included the 1107 bp coding region. The invention
CC relates to the identification and isolation of polynucleotide molecules
CC encoding a new class of protein insecticidal toxins which are produced by
CC bacteria of the genera Xenorhabdus and Photorhabdus (see also AAX07350).
CC The polynucleotide molecules may be incorporated e.g. into insect-
CC specific viruses (e.g. entomopox and nuclear polyhedrosis viruses),
CC bacteria (e.g. Gracilicutes, Firmicutes, Temericutes and Mendosicutes),
CC protozoa, yeast and plants (particularly cereals such as wheat and
CC barley, vegetables such as tomato and potato, fruit trees such as citrus
CC and apple, tobacco and cotton) for control of pest insects. (Updated on
CC 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1107 BP; 424 A; 181 C; 199 G; 303 T; 0 U; 0 Other;
Query Match 65.9%; Score 663.8; DB 2; Length 1107;
Best Local Similarity 79.4%; Pred. No. 2.8e-135;
Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;
QY 12 ATTAACACCTGATGATAGAAAGTATCCACCCGTGAAAGCAATATGCGAGATAT 71
DB 57 AGTAACGGCTGATGATGAAAGGATATCAACCCGTGAAAGCAATATGCGAGATAT 116
QY 72 AGTAGTATCTAATCACTTTAAGCAAAAGATGAGGTCTATACAGCATCATATGAAATGA 131
DB 117 AATAGGTGATCTAGATTCAGCAAAAGCAATATGATCTATACGATTTGATGAAATTC 176
QY 132 ATATGACCTTAAGAAATATATATATAGCTTACGCTTGTAGTGTATTCATATATGT 191
DB 177 ATATGACCTTAAGAAATATATATATAGCATATATGCTTAAAGCGTATTCATATATGT 236
QY 192 ATCTAACTCTTCTGATGATCTATATTAAGAAATTAAGACTGCTGAGAAATTTATCAAGA 251
DB 237 CTCTCAACTCTCAAGAGACTATATTAATAAATGAATTAACAAGGTAGATTTATCAAGA 296
QY 252 ATATATGCTAATCTTTCACTGCACTATTTAGTGAAATGAGTGAATTTCTTAAGA 311
DB 297 ATACATGCTAATCTTTATCTGCACTATTTGAGTATGAGATGATCAATTTCTTAAGA 356
QY 312 TATGCAAAATGCTTTTATATAGAAATGAATCTGATTTTGAAGGTCAATATCTCAAAAAT 371
DB 357 TATGCAAAATGATTTTATCCCAAGAACGATGAGTTTGAAGGTCAACGCTTAAAAATAC 416
QY 372 TTGGAATGTTCTGAGCTTGAATAAATTAACATGAGTCTTATTCAGATGAGATTAAT 431
DB 417 CTGGGATATTCCTGATCTTGAGAAATTAATTTGAAAGTTATTCAGATGAGATTAAT 476
QY 432 ATTAGCAGATATTTTCTCTGTACAGGAATTCACATGAGGAGAAATCAATCAAA 491
DB 477 ATTAGCAGATATTTCTTCTTGTCTTCAAGAACTTCAATGAGGCAATCAATCAAA 536

```

```

QY 492 TGCCGCAAGATTTTAAATTAATTGATTTCTTATTTACCTTATCTGCTGTAACCTTACT 551
DB 537 TGACGCAAAATTTTAAATTAATTGATTTTCTTATTTCTTATCTGCTGTAACCTTACT 596
QY 552 GGGAAAGGAGATTTTCAAAAACCTTTACAAATGATTTAGAGGCTAATATCATTTAGAGA 611
DB 597 GGGAAAAAGATTTTCAAAAATTTTACAAATGCTTACAAATCAATATCATTTAGAGA 656
QY 612 TTATATGAGAAAAAATCTTTCAAACTTTCTTTGACACCGCAGAGATTAACCTGA 671
DB 657 TTATATGAGAAAAAATCTTTCAAACTTTCTTTGACACCGCAGAGATTAACCTGA 716
QY 672 TGCGAAATGAGTTATTTGCTGACCAACGAGAGCGCTTAA----- 714
DB 717 TGCGAAACAGGCTACTTGGCCGGTCAACAAAGCGCTTAAATTTGCCAACACGCTCTTC 776
QY 715 -----TGAGAGTGAAGTTTAAAGACTTAA 740
DB 777 TACAGCAAAACGCTACAGACGCTTCACTTAATTTGAGAGTTAGTTGCAAAAACCTTAG 836
QY 741 AATTAACAAATCTAGAAATGATTTTCTAATATGAGAAAGGCTGCAAAAACAAAGTATAG 800
DB 837 AGATAACCATCCAGAAATACATTTATGAAATGATGATGCTGCAAAAACGAAATATAG 896
QY 801 TTCAATTTTAAAGAGTTACAAAAGGTTAACGCTTCCACAGACGACGCAAAAAGTATTGG 860
DB 897 TTCAATTTTAAAGAGTTACAAAAGGTTAACGCTTCCACAGACGACGCAAAAAGTATTGG 956
QY 861 TACAGCGATGAGGAGTAACTGGAATAATTTGCCGAATA--ATTATATAGTGAAGCT 917
DB 957 TACAAAAGCGGAGTAACTTTCGAAAACCTGCAAGGTAGAGATTTATATATAGTAACT 1016
QY 918 AAGCAAAAAGACAGGGTAACTTTACTCAAAATGATATCTGACATATACATGACCGTTCA 977
DB 1017 AAGCAAAAGACAGGGTAACTTCTCATTAATATATCTGACCAAAATATATGAGATCA 1076
QY 978 TAGTGTGAACTCATTTAATAAATATATGTA 1008
DB 1077 AAGTGTGAACTCATTTACAAAATATATATTA 1107
RESULT 7
AAX07351
ID AAX07351 standard; DNA; 1205 BP.
XX
XX AAX07351;
AC 17-OCT-2003 (revised)
DT 21-MAY-1999 (first entry)
XX
DE Xenorhabdus nematophilus insecticidal toxin gene toxb4.
XX
KW Toxin; toxb4; biological control; insecticide; ds.
XX
OS Xenorhabdus nematophila.
XX
FH Key Location/Qualifiers
FT CDS 17..1123
FT /tag=a
FT /note="the coding region is specifically claimed in
FT Claim 1"
XX
XX WO9903328-A1.
XX
XX 28-JAN-1999.
XX
XX PF 17-JUL-1998; 98WO-AU000562.
XX
XX PR 17-JUL-1997; 97AU-00008088.
XX
XX PA (CSTR ) COMMONWEALTH SCI & IND RES ORG.

```

Db 6.13 GCGAAAAAGGATTTTTCAAAAATTTTACATGGTCTAGAACTAATCATTAGAGAA 672

DR WPI; 1995-052084/07.

XX Smigielski AJ, Akhurst RJ;  
PI  
XX  
DR WPI; 1995-052084/07.  
DR P-PSDB; AAR68961.

xx Polynucleotide(s) that encode insecticidal toxins from *Xenorhabdus* -  
PT useful for the control of insect pests in the agricultural, aquatic and  
XX forest industries.

PS Claim 1, Page 9, 15pp; English.

CC This nucleic acid sequence encodes an insecticidal toxin which is useful  
CC for genetically engineering a wide range of biological systems which will  
CC become more useful for the control of insect pests detrimental to  
CC agricultural, aquatic and forest industries. The sequence may be  
CC expressed in recombinant organisms. (Updated on 25-MAR-2003 to correct PN  
CC field.) (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 1272 BP, 483 A, 211 C, 230 G, 348 T, 0 U, 0 Other;

Query Match 62.3%; Score 628.2; DB 2; Length 1272;  
Best Local Similarity 79.1%; Pred. No. 1.7e-127;  
Matches 833; Conservative 0; Mismatches 163; Indels 57; Gaps 5;

```
QY 12 ATTAACACCTGATGATGAGATGATATCCACCCGTTGAAAACCAATATGACAGATAT 71
DB 139 AGTAACGCTGATGATGATGATATCAACCCGTTGAAAACCAATATGCGGAGATAT 198
QY 72 AGTACGATACCTAACTTAAACCAACAGATGAGGTCATACGATCATATGAAATGA 131
DB 199 AATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 258
QY 132 ATATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191
DB 259 ATATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 318
QY 192 ATCTAACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251
DB 319 CTCTCAACTTCAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 378
QY 252 ATATATGCTATCTTTCATCTGCACTATGATGATGATGATGATGATGATGATGAT 311
DB 379 ATACATGCTATCTTTCATCTGCACTATGATGATGATGATGATGATGATGATGAT 438
QY 312 TATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
DB 439 TATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 497
QY 372 TTGAATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 430
DB 498 CTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 557
QY 431 TATTAAGCACTATCTTCTCTGACAGAAATTCCTGAGAGAAATTCATCATCA 490
DB 558 TATTAAGCACTATCTTCTCTGACAGAAATTCCTGAGAGAAATTCATCATCA 617
QY 491 ATGCGCAAGATTTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 550
DB 618 ATGCGCAAGATTTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 677
QY 551 TGGGAAGAGATTTTTTCAAAAACTTTTACAAATGATGATGATGATGATGATGAT 610
DB 678 TGGGAAGAGATTTTTTCAAAAACTTTTACAAATGATGATGATGATGATGATGAT 737
QY 611 ATTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 670
DB 738 ATTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 797
QY 671 ATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 714
DB 798 ATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 857
QY 715 -----TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
DB 858 CTACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 917
QY 739 AAAAATTAACAATTAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 798
```

```
DB 918 AGAGTAATCCCATCCAGAAATATCATTTATGAAATGATGATGCTGCAAAACGAAATAT 977
QY 799 AGTTCATTTATTAAGAGAGATCAAAAAGGTTAAGCTCCACAGACAGACGAAAGAT 858
DB 978 AGTTCATTTATTAAGAGAGATCAAAAAGGTTAAGCTCCACAGAGAGAGAGAT 1037
QY 859 GGTACAGCAGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 915
DB 1038 GGTACAGCAGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1097
QY 916 CTAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 975
DB 1098 CTAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1157
QY 976 CATAGTGTGGAACCTCATTAATTAATTAATTAATTAATTAATTAATTAATTA 1008
DB 1158 CAAGTGTGGAACCTCATTAATTAATTAATTAATTAATTAATTAATTAATTA 1190
```

RESULT 9  
AAA70101  
ID AAA70101 standard; DNA; 3927 BP.  
XX  
AC AAA70101;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:234.  
XX  
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KM antimalarial; malaria; protozoicide; infection; insecticide; ds.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO20025728-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-US026796.  
XX  
PR 05-NOV-1998; 98US-0107131P.  
XX  
PA (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
PI Hoffman S, Carnucci D, Gardner M, Venter JC;  
XX  
DR WPI; 2000-365347/31.  
XX  
PT Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection.  
XX  
PS Disclosure; Page 458-459; 577pp; English.  
XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II). (1) and  
CC (II) are useful for the development of vaccines against P. falciparum  
CC infection. (1) and polyclonal antisera or a monoclonal antibody raised to  
CC immunogens comprising the sequences of (I), are useful in the detection  
CC of infection with P. falciparum. Furthermore, (I) (especially when they  
CC are refined or secreted or membrane proteins) can aid the identification  
CC of drugs to treat or prevent P. falciparum infection, or they can be used  
CC to identify drug resistance in P. falciparum. Sequencing of the  
CC Plasmodium chromosome 2 and the subsequent identification of proteins  
CC encoded by it will help to expand our understanding of parasite biology,  
CC a process hampered by the complexity of the parasitic lifecycle, and  
CC provide new targets for vaccine and drug development. Parasite resistance



QY 441 ATATTTTCTCTGACAGAAATTCACGAGGAGAAATCAACAATCAATGCCGCAAG 500  
DB 1068 TTTTATTTATTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1009  
QY 501 ATTTTAAATTAATGATTTCTTTATTTCTTATCTGCTGTAATCTTCACTGGGAAGAG 560  
DB 1008 ATTTTATTTTAAATAATTAATTAATTAATTTTATTTATTTATTTATTTATTTATTTAAT 949  
QY 561 GATTTTTCACAAAACCTTTTACATGATGATGAGGCTAAATCATTTAGAGATTTATTTGA 620  
DB 948 TATTTTAAAAAAAATTTTATCAAAAAACAACAAAATTAATTAATTAATTAATTAATTA 889  
QY 621 GAGAAAAAACCTTTCTTAA 639  
DB 888 ATTACAAATTTTATACA 870

## RESULT 11

AA60472  
ID AA60472 standard; DNA; 4590 BP.

XX AA60472;

XX 25-MAR-2003 (revised)

DT 24-AUG-1991 (first entry)

XX Sequence encoding the ring-infected Erythrocyte Surface Antigen (RESA).

KW Malaria vaccine; antigen; epitope; ss.

XX Plasmodium falciparum.

XX Key Location/Qualifiers

FT exon 801..995

FT /\*tag= a

FT 1199..4225

FT /\*tag= b

PN W08601802-A.

XX 27-MAR-1986.

XX 11-SEP-1984; 84AU-00007066.

XX 11-SEP-1984; 84AU-00007066.

XX 10-SEP-1985; 85AU-00047326.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;

XX WPI; 1986-094065/14.

XX P-PSDB; AAF60569.

XX DNA coding for Plasmodium falciparum antigens - expressing

XX poly:peptide(s) having antigenicity of RESA or FIRA antigens of P

XX falciparum.

XX Claim 4; Fig 1; 55pp; English.

XX The inventors claim a novel DNA molecule which comprises a nucleotide

XX sequence corresp. to all or a portion of the base sequence coding RESA

XX (AA60472) or FIRA (AA60473). RESA and FIRA have antigenicity suitable

XX for providing protective immunity against Plasmodium falciparum malarial

XX infections. (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T; 0 U; 0 Other;

XX Query Match 5.0%; Score 50.8; DB 1; Length 4590;

XX Best Local Similarity 43.6%; Pred. No. 0.24;

XX Matches 226; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 1 ATGTTATACATTTATACACTGATGATGAAAGTGATATCCACCCGTTGAAAAGCAATTA 60  
DB 3587 ATTTGAGATATATATATTCATGAAATTCAGCAATTAATTAACAAATTAATTAATTAAT 3646  
QY 61 GCAGAGAT 120  
DB 3647 GAAATATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3706  
QY 121 TATGAAATGAAAT 180  
DB 3707 GAAATATGAAAT 3766  
QY 181 ATTCATTAATGAT 240  
DB 3767 CATGATGCTGAAAGAAATGTAAGAAATGTAAGAAATGTAAGAAATGTAAGAAATGTAAG 3826  
QY 241 ATTTATCAAGAT 300  
DB 3827 GAAATATGAAAT 3886  
QY 301 ATTTCTAAAGAT 360  
DB 3887 GTAGAAAGAAATGTAAGAAATGTAAGAAATGTAAGAAATGTAAGAAATGTAAGAAAT 3946  
QY 361 CCTCAAAACATTTGGAATGTTCTGAGCTTGAATTAACCAATTAAGATGCTTATTCAGAT 420  
DB 3947 GTTGAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGAAGAAAT 4006  
QY 421 GACGATTAAT 480  
DB 4007 GATGAGAAATATGTTGAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGAAGAAAT 4066  
QY 481 CAACATCAAAATGCCGCAAGATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 518  
DB 4067 GAAAGAAATGTAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGAAG 4104

## RESULT 12

ADL08108  
ID ADL08108 standard; DNA; 188971 BP.

XX ADL08108;

XX 20-MAY-2004 (first entry)

XX Human gene associated with low HDL-C APOA1.

XX Human; ds; SNP; single nucleotide polymorphism;

XX high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;

XX coronary artery disease; gene.

XX Homo sapiens.

XX Key Location/Qualifiers

FT variation replace(123408,T)

FT /\*tag= a

FT /standard\_name= "Single nucleotide polymorphism"

PN US2004043389-A1.

XX 04-MAR-2004.

XX 04-SEP-2002; 2002US-00235192.

XX 04-SEP-2002; 2002US-00235192.

XX (VITTI-) VITTI VITY INC.

XX McCarthy J;

XX WPI; 2004-214170/20.

XX Determining whether a subject has, or is at risk of developing, an







ID AAA70105 standard; DNA; 5940 BP.  
 AC AAA70105;  
 XX  
 DT 07-NOV-2000 (first entry)  
 XX  
 DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:238.  
 XX  
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
 XX antimalarial; malaria; protozoacide; infection; insecticide; ds.  
 OS Plasmodium falciparum.  
 XX MO200025728-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 05-NOV-1999; 99WO-US026796.  
 XX  
 PR 05-NOV-1998; 98US-0107131P.  
 XX  
 PA (HOFF/) HOFFMAN S.  
 XX (CARU/) CARUCCI D.  
 PA (GARD/) GARDNER M.  
 PA (VENT/) VENTER J C.  
 XX  
 PI Hoffman S, Carucci D, Gardner M, Venter JC;  
 XX  
 DR WPI: 2000-365347/31.  
 XX  
 PT Proteins encoded by chromosome 2 of the human malarial parasite,  
 XX Plasmodium falciparum, useful as antimalarial vaccines and in the  
 PT diagnosis of P. falciparum infection.  
 XX  
 PS Disclosure: Page 460-462; 577pp; English.  
 XX  
 CC The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
 CC vaccines against P. falciparum infection comprising (i) or (ii). (i) and  
 CC (ii) are useful for the development of vaccines against P. falciparum  
 CC infection. (i) and polyclonal antisera or a monoclonal antibody raised to  
 CC immunogens comprising the sequences of (i), are useful in the detection  
 CC of infection with P. falciparum. Furthermore, (ii) (especially when they  
 CC are refined or secreted or membrane proteins) can aid the identification  
 CC of drugs to treat or prevent P. falciparum infection, or they can be used  
 CC to identify drug resistance in P. falciparum. Sequencing of the  
 CC Plasmodium chromosome 2 and the subsequent identification of proteins  
 CC encoded by it will help to expand our understanding of parasite biology,  
 CC a process hampered by the complexity of the parasitic lifecycle, and  
 CC provide new targets for vaccine and drug development. Parasite resistance  
 CC to drugs and mosquito resistance to insecticides have led to a resurgence  
 CC of malaria in many parts of the world, and there is a pressing need for  
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352  
 CC represent nucleotide and protein sequences given in the present  
 CC invention, but which are not specifically mentioned within the  
 CC specification  
 CC  
 SQ Sequence 5940 BP; 3106 A; 343 C; 879 G; 1612 T; 0 U; 0 Other;

Query Match 4.9%; Score 49.2; DB 3; Length 5940;  
 Best Local Similarity 43.3%; Pred. No. 0.56;  
 Matches 284; Conservative 0; Mismatches 368; Indels 4; Gaps 1;

QY 175 AGTGTATTCATATGATCTTAACCTTCCTGATGACTATTTATTAAGATTAAGAGACTGCT 234  
 DB 229 AATATATATGATTAATTAATGATTAACAATATGATTAATTAATGATTAATATGAT 288  
 QY 235 GAGAGATTTATCAAGATATATGCTATCTTCATCTGCACCTATTTAGTGAANAATGCT 294  
 DB 289 AATTAATTAATGATTAATTAATGATTAATGATTAATTAATTAATTAATTAATGATC 348  
 QY 295 GATCAAAATTTCTAAGATATGCAAAATGCTTTTATTAAGATGAATGATTTTGAAGCT 354

DB 349 GATGAATATCAAAAAATATATATACATTAAGACATGAGTGAAGAAACACCTTAAGAT 408  
 QY 355 CAATATCCTCAAAACATTTGGAATGTTCCGAGCTGAAAATTAACCATGAGCTTAT 414  
 DB 409 ACATTAAGATCCATTAGTTCGTTGCGAATTAATGGAATTAAGAAATTAATGA 468  
 QY 415 TCAGATGACGATTAATTAATTAATGACTATATTTTCTGTACAGAAATTCACCTGAG 474  
 DB 469 GAATTTAGAAAAAGATTTAAAGAGTAAGATTAAGATTAATTAATTAATTAATGA 528  
 QY 475 GAAATCAACATCAATCAATGCGCAGATTTTAAATTAATTAATTAATTAATTAATCACTTA 534  
 DB 529 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 584  
 QY 535 TCTGCTGTAATTCACCTGGAAGAGATTTTTCAAAACCTTTTAAATGAATTAAG 594  
 DB 585 AAAAGAAAATCTTTTACAAAGAAAAGATTAATTAATTAATTAATTAATTAATTAATGA 644  
 QY 595 GCTAATCATTAAGAAATTAATTAATTAAGAAAGAAACCTTTCTTAAACCTTTCTGACCA 654  
 DB 645 TGAAAAAGAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 704  
 QY 655 CCGAGAGATTAACCTGATGCGAAGATTAATTTGCTGACCAAGAAAGCGCTTAA 714  
 DB 705 TTTAGAAAAATTAATTAAGAAAGAAACGATTTCTATAGAAATTAATTAATTAATTAATGA 764  
 QY 715 TGGAGATGATTTTAAAGAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 774  
 DB 765 TCTAGAAAAATTAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATGA 824  
 QY 775 GAAAGGCTGCAAAACAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 830  
 DB 825 AAAATTTATCAAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 880

Search completed: November 21, 2004, 06:27:15  
 Job time : 546 secs

THIS PAGE BLANK (USPTO)

THIS PAGE

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_nzp model

Run on: November 21, 2004, 08:42:25 ; Search time 29.5 Seconds

(without alignments)  
4532.105 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1735

Sequence: 1 atggttatcaattatacacc.....ccctattataaataatataga 1008

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 478139 segs, 66318000 residues

Total number of hits satisfying chosen parameters: 956278

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_nzp.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.epool.p/US10617962/rnat.17112004.163923.9771/app.query.fasta.1.1159  
-DB=Issued Patents AA -QFMT=fastan -SUPFIX=rai -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10617962.0CGN\_1\_1\_33.0/rnat.17112004.163923.9771 -NCPU=6 -ICPU=3  
-NO MAP -LARGEDUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1734	99.9	335	4	US-09-463-048A-4
2	1276	73.5	368	4	US-09-463-048A-3
3	1058	61.0	383	2	US-08-569-168-7
4	940.5	54.2	278	2	US-08-569-168-2
5	105.5	6.1	816	4	US-09-328-352-6677
6	101.5	5.9	600	4	US-09-248-795A-23971
7	96.5	5.6	293	1	US-08-628-291-4
8	96.5	5.6	293	2	US-09-128-722-4
9	96.5	5.6	317	1	US-08-628-291-12
10	96.5	5.6	317	2	US-09-128-722-12
11	96	5.5	623	4	US-09-538-092-119
12	96	5.5	912	2	US-08-951-871-2

13	96	5.5	3168	4	US-09-489-039A-14067	Sequence 14067, A
14	95.5	5.5	680	4	US-09-583-110-3191	Sequence 3191, Ap
15	95.5	5.5	915	1	US-08-328-322-5	Sequence 5, Appl1
16	94.5	5.4	400	4	US-09-248-796A-16641	Sequence 16641, A
17	94.5	5.4	696	3	US-08-899-437-23	Sequence 23, Appl
18	94.5	5.4	696	3	US-09-126-121-23	Sequence 23, Appl
19	94.5	5.4	753	4	US-09-543-681A-5022	Sequence 5022, Ap
20	94.5	5.4	863	2	US-08-666-271-2	Sequence 5022, Ap
21	94	5.4	325	4	US-09-134-000C-4195	Sequence 4195, Ap
22	94	5.4	534	4	US-09-538-092-673	Sequence 673, App
23	94	5.4	910	4	US-09-623-326-7	Sequence 673, App
24	94	5.4	1388	2	US-08-685-576-1	Sequence 1, Appl1
25	92.5	5.3	688	4	US-09-248-796A-15231	Sequence 15231, A
26	92.5	5.3	1572	2	US-08-290-731C-5	Sequence 5, Appl1
27	92.5	5.3	1596	3	US-09-356-952-3	Sequence 3, Appl1
28	92	5.3	598	2	US-08-937-540-2	Sequence 2, Appl1
29	92	5.3	598	4	US-09-398-395A-26	Sequence 26, Appl
30	92	5.3	598	4	US-09-887-586A-26	Sequence 26, Appl
31	92	5.3	598	4	US-09-895-752-26	Sequence 26, Appl
32	92	5.3	598	4	US-09-903-012B-26	Sequence 26, Appl
33	92	5.3	598	4	US-09-900-797-26	Sequence 26, Appl
34	91.5	5.3	399	4	US-09-107-532A-4714	Sequence 4714, Ap
35	91.5	5.3	475	4	US-09-248-796A-17531	Sequence 17531, A
36	91.5	5.3	485	4	US-09-639-245-8	Sequence 8, Appl1
37	91	5.2	311	4	US-09-248-796A-16692	Sequence 16692, A
38	91	5.2	910	4	US-09-623-326-8	Sequence 8, Appl1
39	90.5	5.2	393	4	US-09-543-681A-6487	Sequence 6487, Ap
40	90	5.2	415	4	US-09-543-681A-6746	Sequence 6746, Ap
41	90	5.2	447	3	US-08-961-083-182	Sequence 182, App
42	90	5.2	447	4	US-09-536-784-182	Sequence 182, App
43	90	5.2	484	4	US-09-468-656A-6	Sequence 6, Appl1
44	90	5.2	1039	4	US-09-583-110-5226	Sequence 5226, Ap
45	90	5.2	1388	2	US-08-685-576-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-463-048A-4  
; Sequence 4, Application US/09463048A  
; Patent No. 6630619  
; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
; APPLICANT: EAST, Peter David  
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photob  
; FILE REFERENCE: 050179-0076  
; CURRENT APPLICATION NUMBER: US/09/463,048A  
; CURRENT FILING DATE: 2002-12-13  
; PRIOR APPLICATION NUMBER: PCT/AU98/00562  
; PRIOR FILING DATE: 1998-07-17  
; PRIOR APPLICATION NUMBER: PO 8088  
; PRIOR FILING DATE: 1997-07-17  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 4  
; LENGTH: 335  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-463-048A-4

Alignment Scores:  
Pred. No.: 2.5e-174  
Score: 1734.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 99.94%  
DB: 4  
Gaps: 0

US-10-617-962-2 (1-1008) x US-09-463-048A-4 (1-335)

QY 1 ATGTTATCAATTAAACCTGATGATAGAGTATATCCACCGTTGAAAGCAATA 60  
|||||

Db 1 MetValIIeGlnLeuThrProAspAspArgSergIlyTyrrProProValGluIysGlnIle 20  
Qy 61 GCAGAGATATAGTACGTATCTAACTTTAAGCAACAGATGAGGTCATACAGCATCA 120  
Db 21 ALaGlyAspIleValArgIleLeuAsnPhelyGlnThrAspGluGlyHisThrIleSer 40  
Qy 121 TATGGAATTTGAATATGAGCTAAGAAATTAATTTAGCTTACGCTTTGGCTGAATGGT 180  
Db 41 TyrgIyIleGluIyrrArgAlaIySvlIleIleuAlaIyrrAlaIleuAlaValSergIy 60  
Qy 181 ATTCAATATGATCTAACTTCTCGATGACTATTTATAGAAATTAAGAGACTGCTGAGAG 240  
Db 61 ILeHisAsnValSerIySerIeuProAspAspTyrrIlyAsnIySgluThrIleGluIy 80  
Qy 241 ATTATCAAGATATATATGTCATATCTTTCATCTGCACTATTTAGGTGAAAATGGTGTCA 300  
Db 81 ILeTyrrGlnIyrrIyrrMetSerIeuAsnIeuSerSerAlaIleuIeuGlyIuAsnGlyIy 100  
Qy 301 ATTTCTAAAGATATGCGCAATGGTTTTTATNAGAATGAACTGGATTTTGAAGTCAATAT 360  
Db 101 ILeSerIyAspIleValArgIleAsnGlyPheTyrrIyAsnGluIeuAspPheGluGlyIy 120  
Qy 361 CCTCAAAACATTTGGAATGTTCTCGAGCTTGAATAATTAACCATTTGAGTGTATTCAGAT 420  
Db 121 ProGlnAsnIleTyrrAsnValProGluIeuGluAsnIySProIeuSerAlaIyrrSerAsp 140  
Qy 421 GACGATTAATTTATGACCTATATATTTTCTGTGACAGAAATTCACATGAGAAAT 480  
Db 141 AspAspIySerIeuAlaIeuAlaIyrrPhePheSerValGlnIyrrIleProIeuGlnIy 160  
Qy 481 CAACATCAATGCGCCAGATTTTAAATTTTAAATTAATGATTTCTATTTACCTTATCTGCT 540  
Db 161 GlnGlnSerIyAsnIleAlaIyrrPhePheIySLeuIleAspPheIyrrIleuSerAla 180  
Qy 541 GTAATCTCACTGGGAGAGAGATTTTTCAAAAAACTTTTACATGATGATGAGGCTAA 600  
Db 181 ValThrSerIeuGluIyrrArgIlePheSerIyAsnPhelyrrAsnGlyIeuGluIyAla 200  
Qy 601 TCATTNAGAGATATATATGAGAGAAAAAATTTTCTTAACTTTTTCGACCAACCGCAG 660  
Db 201 SerIeuGluIyAsnTyrrIleGluIyrrGlySvlIySLeuSerIySProPheAspProGln 220  
Qy 661 AGATTACCTGATGCGAGAAATGTTTATTTGGCTGCGACCAACAGAAACGCTTAATGANA 720  
Db 221 ArgIeuProAspIyrrArgIleGlyTyrrIleuAlaGlyProThrIyrrIleuAlaProIy 240  
Qy 721 GTGAGTTTAAAGAACTTAAATAATTAACAATCTAGAAATGATTTTCTAATATGAAAGG 780  
Db 241 ValSerPheIySgluIeuIySAsnAsnIySerIyrrArgAsnIyPheSerIyrrMetGlu 260  
Qy 781 GCTGCAAAACAAAGATATGTTTATTTATTAAGAGGTACAAAGGGTACCGTCCACAG 840  
Db 261 AlaAlaIySgluIyrrIyrrSerSerPheIleIySgluValGlnIySgluAsnAlaProGln 280  
Qy 841 ACAGGACCCGAAAGATTTGGTACAGCCAGTGGCAGTAACTGGAAAAATTTGCCGAATAT 900  
Db 281 ThrAlaAlaIySerIyrrIleGlyThrAlaSerGlySerIyrrAsnIySgluIySLeuProAsn 300  
Qy 901 TTATATAGTGTGAGGCTTAAGCCAAAAAGACAGGTTAACTTTTCTCAAAATGATTAAGTAC 960  
Db 301 LeuTyrrSerValArgIeuSerGlnIySAspArgValThrPheThrGlnIyrrAsnPhel 320  
Qy 961 AATACATGACGGTTCATAGTGTGGAATCTCATTTATTAATAATAATA 1005  
Db 321 AsnThrMetThrValHisSerValGlyThrHisTyrrIySAsnIle 335

RESULT 2  
US-09-463-048A-3  
; Sequence 3, Application US/09463048A  
; Patent No. 6630619  
; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
; APPLICANT: EAST, Peter David

; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photoc  
; FILE REFERENCE: Luminescens  
; CURRENT APPLICATION NUMBER: US/09/463,048A  
; CURRENT FILING DATE: 2002-12-13  
; PRIOR APPLICATION NUMBER: PCT/AU98/00562  
; PRIOR FILING DATE: 1998-07-17  
; PRIOR APPLICATION NUMBER: PO 8088  
; PRIOR FILING DATE: 1997-07-17  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 3  
; LENGTH: 368  
; TYPE: ERT  
; ORGANISM: Xenorhabdus nematophilus  
US-09-463-048A-3

Alignment Scores:  
Pred. No.: 5,13e-126 Length: 368  
Score: 1276.00 Matches: 251  
Percent Similarity: 80.17% Conservative: 36  
Best Local Similarity: 70.11% Mismatches: 47  
Query Match: 73.54% Indels: 24  
DB: 4 Gaps: 3

US-10-617-962-2 (1-1008) x US-09-463-048A-3 (1-368)

Qy 4 GTTATACATTAATTAACCTGATGATAGA-----AGTGAATATCCACCC 45  
Db 11 ValIIeGlnLeuThrProAspAspArgValThrProAspAspIySgluIyrrGlnPro 30  
Qy 46 GTTGAAGCAAAATATGACAGATATAGTATGATTAATTAACCTTTAAGCAACAGATGAG 105  
Db 31 ValGluIySgluIleAlaIyrrIleIleArgValIleuGluPheIySgluIyrrAsnGln 50  
Qy 106 GGTATACAGCATCATATGGAATTTGAATTTGAGGCTAAGAAATATATAGCTTACGCT 165  
Db 51 SerHisThrGlyLeuTyrrIyrrIleAlaIyrrArgAlaIySvlIleIleIleAlaIyrrAla 70  
Qy 166 TTGGCTGATGAGTGAATTCATATGATGATTAACCTTCTGATGATATTAAGATATA 225  
Db 71 LeuAlaValSerGlyIleHisAsnValSerGlnIeuProGluAspTyrrIyrrAsnIyS 90  
Qy 226 GAGACTGCTGAAGAAATTTATCAAGATATATGCTTAATCTTTCATCTGCACTATTAAGT 285  
Db 91 AspAsnThrGlyrrArgIleTyrrGlnIyrrIyrrMetSerIyrrAsnIeuIeuSerIyrrAlaIeuIeuGly 110  
Qy 286 GAAATATGATGATCAATTTCTAAGATATGGAATGTTCTGAGCTTGAATAATTAACCATTTG 345  
Db 111 GluAsnGlyAspGlnIleSerIyrrAspPheAlaAsnAspPheThrGlnAsnGluIeuGln 130  
Qy 346 TTTGAAGTCAATATTCCTCAAAACATTTGGAATGTTCTGAGCTTGAATAATTAACCATTTG 405  
Db 131 PheGlyGlyGlnIyrrArgIeuIyrrAsnThrIyrrAspIleProAspIeuGluIyrrIyrrAsnIyS 150  
Qy 406 AGTGTATTTACATGATGACGATTAATTTATTAAGCACTATATTTTCTCTGTAAGAAAT 465  
Db 151 GluAspTyrrSerAspIyrrAspIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrr 170  
Qy 466 CCACTGAGGAAATATCAATCAATCAATGCGGAAGATTTTAAATTAATTAAGTTCCTTA 525  
Db 171 PrometGluAlaAsnIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrr 190  
Qy 526 TTTACCTTATTCGCTGATTAATTCACCTGAGAGAGAGATTTTTCAAAACTTTTACAT 585  
Db 191 LeuIleIeuSerAlaValThrSerIeuGlyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrr 210  
Qy 586 GGATTNAGAGCTTAATCATTAAGATTAATTTATTAAGAGAAAAAATTTCTTAACTTTC 645  
Db 211 GlyIeuGluIyrrIyrrSerIeuGluIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrrIyrr 230  
Qy 646 TTTGACCAACCGGAGAGATTAATCTGATGAGCAAAATAGGTTATTTGGCTGACCAACAGAA 705

Db 231 pheargProProGlnInlyLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrIlys 250  
 QY 706 GGGCCATAA----- 714  
 Db 251 AlaProlyLeuProThrThrsSerSerThrAlaThrSerThrAlaAlaSerSerAsn 270  
 QY 715 TCGAGAGTGAAGTTTAAAGACTTAAATAAATAAATACTAGATGATGATTTCTAATATG 774  
 Db 271 TPAAGValSerLeuGlnInlyLeuAlaGAspAsnProSerAlaGAsnThrhehecllySmet 290  
 QY 775 GAAGGGCGCTGCAAAACAAAGTATGCTTATTATTAAGAGGTACAAAGGGTAAAGCT 834  
 Db 291 AspaSAlaAlaLysArgLysTyrSerSerPheIleLysGluValGlnInlySgIyAsnAsp 310  
 QY 835 CCAACAGACGACGCGAAAGTATGTTGTAAGCAGCGAGTAACTGGAAATAATTGGCCG 894  
 Db 311 ProAlaAlaAlaAlaAlaSerIleGlyThrLysSerGlySerAsnheGlnInlySgIy 330  
 QY 895 ---AATTAATTTATATGTTGAGGCTAAGCCAAAGAGAGGTAACCTTTACTCAAAAT 951  
 Db 331 GlyArgAspLeuTyrSerIleArgLeuSerGlnGlnInlyAlaGValThrheSerIleAsn 350  
 QY 952 GATACTGACAAATACATGACGGCTTCACTGTTGAACTCATTAATAATAATA 1005  
 Db 351 AsnThrAspGlnIleMetGluIleGlnSerValGlyThrHisTyrGlnAsnIle 368

## RESULT 3

US-08-569-168-7  
 / Sequence 7, Application US/08569168  
 / Patent No. 5972687  
 / GENERAL INFORMATION:  
 / APPLICANT: SmigielSKI, Adam J.  
 / TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS  
 / NUMBER OF SEQUENCES: 7  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Lowe, Price, Leblanc & Becker  
 / STREET: 99 Canal Center Plaza, Suite 300  
 / CITY: Alexandria  
 / STATE: VA  
 / COUNTRY: USA  
 / ZIP: 22314  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / OPERATING SYSTEM: IBM PC compatible  
 / SOFTWARE: PatentIn Release #1.0, Version #1.30  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/569,168  
 / FILING DATE:  
 / CLASSIFICATION: 424  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Price, Robert L.  
 / REGISTRATION NUMBER: 22,685  
 / REFERENCE/DOCKET NUMBER: 1451-015  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 703-684-1111  
 / INFORMATION FOR SEQ ID NO: 7:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 383 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-08-569-168-7

## Alignment Scores:

Pred. No.: 5,06e-103  
 Score: 1058.00  
 Percent Similarity: 73.688  
 Best Local Similarity: 65.108  
 Query Match: 60.984  
 DB: 2  
 Gaps: 10

US-10-617-962-2 (1-1008) x US-08-569-168-7 (1-383)

QY 4 GTTATACAAATTAAACCCGATGATGA-----AGTGATATCCACC 45  
 Db 11 ValIleGlnLeuThrProAspAspArgValThrProAspAspLysGlyGluIyrgIlnPro 30  
 QY 46 GTTGAAGAACAAATAGCAGAGATATAGCTATTAATACTTAAAGCAACAGATGAG 105  
 Db 31 ValGluIySgIlnIleAlaGlyAspIleIleArgValIleuGlnPheLysGlnThrAsnGln 50  
 QY 106 GGTCTATACAGCATCATATGAAATGAAATATGAGCTTAAGAAATAATATATTAAGCTTACGCT 165  
 Db 51 SerHisThrGlyLeuTyrGlyIleProIyrgArgAlaLysValIleIleAlaTyrAla 70  
 QY 166 TTGGCTGTAAAGTGATTCATTAATGATTAATAACTCTGATGACTATTAATAAGATAA 225  
 Db 71 LeuAlaValSerGlyIleHisAsnValSerGlnLeuProGluAspIyryIyryAsnIyS 90  
 QY 226 GAGACTGCTGAGAGAAATTTATCAAGAAATATATGCTAATCTTCACTGCACTTAATAGGT 285  
 Db 91 AspaSntThrGlyArgIleTyrGlnValTyrMetSerAsnLeuSerAlaLeuLeuGly 110  
 QY 286 GAAATATGTCATCAATTTCTAAGATATGCGAAATGCTTTTATAGAAATGACTGAT 345  
 Db 111 GluAsnGlyAspGlnIleSerIySAspMetAlaAsnAspPheThrGlnAsnGluLeuGln 130  
 QY 346 TTTGAAGTCATATCTTCATAAACAATTGGAAATGCTTCTGAGCTTGAAATAATACCATG 405  
 Db 131 PheGln-ValAsnValIleuLysIleProGlyIlePheLeuIleuAlaArgIleAsnIyryr 150  
 QY 406 -AGTCTTATTCAGATGACGATTAATTTATAGCACTATATTTTCTGCTGACGAGAAAT 464  
 Db 150 pLysIleTyrSerAspGluAspLysLeuLeuAlaLeuTyrPhePheAlaSerGlnIle 170  
 QY 465 TCCACTGAGGAAATCAACATCAATCAATGCCGCAAGATTTTAAATTAATTAATGATTTCT 524  
 Db 170 uprMetGluAlaAsnGlnInlySerAlaAlaAsnPhePheLysValIleAspPhele 190  
 QY 525 ATTTACTTATCTGCTGATCACTGAGAGAGATTTTTCAAAAAATTTCATCA 584  
 Db 190 ILeuIleLeuSerAlaValIleThrSerLeuGlyLysArgIlePheSerLysAsnPheTyrAs 210  
 QY 585 TGGATTAGAGCTTAATCATTAAGAAATTAATTAAGGAAATAAATCTTCAACCTT 644  
 Db 210 nGlyLeuGlnThrLysSerLeuGlnAsnTyrIleGluArgLysLysLeuSerLysProPh 230  
 QY 645 CTTTGACCAACCGCAGAGATTAACCTGATGCGAAGATAGGTTATTTGGCTGACCAACAGA 704  
 Db 230 ePheArgProProGlnInlyLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrIly 250  
 QY 705 AGCGCTTAA----- 714  
 Db 250 salAlaProlyLeuProThrThrsSerSerThrAlaThrThrsSerThrAlaAlaSerSerAs 270  
 QY 715 -TGGAGAGTGAAGTTTAAAGAACTTAATAAATAAATACTAGATGATTTCTAATAT 773  
 Db 270 nTPArgValSer-----LeuGlnLysProArgProIleGlnLysTyrIle--Tyr 286  
 QY 774 GGAAGGGGC---TGCAAAACAAAGATATAGTTCATTTTAAAGAGGTACAAAGGGTAA 830  
 Db 287 GluAsnGlyCySgIySerThrIlySile---PheIleTyrLysArgGlyThrIlySgIy--- 304  
 QY 831 CGCTCCACAGACAGACGCAAGAAAGTATGTTGTAAGCAGCGAGTGAAGCTTAAAGAAAT 890  
 Db 305 ---SerThrCySgSerSerSerLysTyrIyryLysLysArgGln--LeuArgLysThr 322  
 QY 891 GCCGAATATATTATATGAGTGAAGCTTAAGCAAAAGAGCAGGTAACCTTATTAACAAA 950  
 Db 323 AlaArgAspPheIle---TyrIlySthrLysProArgThrGlnInlyAsnIleLeuHisLys 341  
 QY 951 TGAATCTGACAAATACATGACGTTTCATATGTTGAACTCATTAATAATAATATATATG 1007

```

Db      34- ---TyPProaSaAbnslYasPrProlys---CysTrPaasSerLeuProLySTyTlle 358

RESULT 4
US-08-569-168-2
/ Sequence 2, Application US/08569168
/ Patent No. 5972687
/
GENERAL INFORMATION:
/ APPLICANT: SmigielSKI, Adam J.
/ APPLICANT: Akhurst, Raymond J.
/ TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lowe, Price, Leblanc & Becker
/ STREET: 99 Canal Center Plaza, Suite 300
/ CITY: Alexandria
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22314
/
COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/569,168
/ FILING DATE:
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Price, Robert L.
/ REGISTRATION NUMBER: 22,685
/ REFERENCE/DOCKET NUMBER: 1451-015
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 703-684-1111
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 278 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
US-08-569-168-2

Alignment Scores:
Pred. No.: 1,11e-90 Length: 278
Score: 940.50 Matches: 191
Percent Similarity: 80.22% Conservative: 24
Best Local Similarity: 71.27% Mismatches: 29
Query Match: 54.21% Indels: 25
DB: 2 Gaps: 2

US-10-617-962-2 (1-1008) x US-08-569-168-2 (1-278)

OY 4 GTTATTCAAATTACACCTGATGATGA-----AGTGATATCCACC 45
Db 11 ValIleGInLeuThrProaSpasPrPaLyValThrProaSpasPrPaLyValGluTyGlnPro 30
OY 46 GTTGAAGAAGCAATATACAGAGATATAGTACGTATCTTAAGCAATTAAGCAAGATGAG 105
Db 31 ValGluLysGlnIleLeaGlyAspIleIleArgValIleGluGlnPheLysGlnThrAsnGlu 50
OY 106 GGTGATACAGACATCATATGGAATTGAAATTCAGCTTAAGAAATAATATATAGCTTACGCT 165
Db 51 SerHisThrClyLeuTyrglyIleProTyrgAlaLysValIleIleAlaTyAla 70
OY 166 TTGGCTGTAGTGGTATTCATATATGTAATCTAACTTCTGATGACTATTAAGAAATTA 225
Db 71 LeuAlaValSerIlyIleHisAsnValSerGlnLeuProGluAspTyTyTlyAsnLys 90
OY 226 GAGACGTGCTGAGAAATTTATCAAGAATATATGTCTTAATCTTCACTTGACATTTAGGT 283
Db 91 AspAsnThrClyArgIleTyrgInValIlyrMetSerAsnLeuLeuSerAlaLeuGly 110
OY 286 GAAATGTGTGATCAAAATTTCTAAAGATATGCGAAATGCTTTTATTAAGATGAATGAT 345

```

```

Db      111 G|uhsncl|yaspdlnl|leSerlyshapMctAlaasnPhethGlnahsncluleuGu 130
QY      346 TTTGAAGTCATATACCTCAAAAATTGGAAATGTTCCTGAGCCTTGAAAAATAAACCATG 405
Db      131 Pheglu-ValasnValleuyleIeProglyIlePheleuileuArgylleAsntyr 150
QY      406 -AAGCTTATTCAGAGACGATTAATTATTACACATATTTTTTCTCGTACAGAAAT 464
Db      150 p|y|l|e|y|S|e|r|S|g|l|a|s|p|l|y|s|e|u|e|u|a|l|e|u|y|r|P|e|h|e|a|S|e|r|g|l|u|e 170
QY      465 TTCACGTGAGGAAATCAACAATCAAAATGCCGCAAGATTTTAAATTAATGATTTCTT 524
Db      170 upromEGlAlaasngclngInserAmalalaasnphelysvalilaspPhee 190
QY      525 ATTACCTTATCTGCTGTACTTCATCTGCAGGAAGAGATTTTTCAAAAAATTTCACA 584
Db      190 u|e|u|i|l|e|u|e|r|S|e|r|a|l|v|a|l|ThrSerleucllysArgliePheSerlysaenPhetyAs 210
QY      585 TGGAATGAGGCTTAAATTCATTAGACATTAATTGAGAGAAAAAATTTCCTTAACTTT 644
Db      210 ngl|y|e|u|g|l|u|Th|y|S|e|r|l|e|u|g|l|u|a|s|n|t|y|r|l|e|g|l|u|a|g|l|y|S|e|u|S|e|r|y|s|P|o|h 230
QY      645 CTTTCGACCAACGCGACAGATTAACCTGATGGACGAATAGGTATTTTGGCTGACCAACAGA 704
Db      230 ePe|h|a|g|P|ro|f|o|g|l|u|s|e|u|P|ro|a|s|p|l|y|a|g|h|r|g|l|y|r|l|e|u|a|l|g|l|P|ro|Th|y 250
QY      705 AGCCCTTAA----- 714
Db      250 sAlaP|o|l|y|S|e|u|P|ro|Th|r|Th|S|e|r|Th|a|Th|r|Th|S|e|r|Th|a|a|l|a|S|e|r|S|e|r|a 270
QY      715 -TGGAGAGTAGCTTTAAAGAA 735
Db      270 ntP|a|r|g|V|a|l|Ser|l|e|u|g|l|n|l|y|s 277

RESULT 5
US-09-328-352-6677
; Sequence 6677, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GT099-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6677
; LENGTH: 816
; TYPE: PRN
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6677

Alignment Scores:
Pred. No.:          0.0184           Length:         816
Score:              105.50           Matches:          96
Percent Similarity: 30.29%           Conservative:     40
Best local Similarity: 21.38%        Mismatches:      128
Query Match:        6.08%            Indels:          185
DB:                 4                Gaps:           22

US-10-617-962-2 (1-1008) x US-09-328-352-6677 (1-816)
QY      112 ACAGCATTATGAAATGTAATGACGTAAAGAAATAATTAAGCT-----TAC 162
Db      216 Th|Se|r|a|l|a|n|S|c|l|n|Th|r|Th|r|V|a|A|g|l|y|a|l|l|e|Th|r|a|l|a|S|p|r|y|r|a|g|T|y|r 235
QY      163 GCTTTGGCTGTAGTGTGTATTCATATATGATCTTAAACTTCGATGACATATTATAAGAT 222
Db      236 Al|a|n|S|c|l|y|P|he|S|e|r|g|l|y|P|he|r|y|-----V|a|l|g|n|Th|r|P|ro|a|S|p|h|r|-----L|y|S|a|l|a 251
QY      223 AAAGAGACTGCTGAGAGAAATTTATCAAGAAATATATGCTTAATCTTCACTGCACATATTA 282

```

Db 252 ArgAlaAsnValSerAsnAlaIlePheValTyrIleProAsnSerSerAlaValIysGly 271  
 QY 283 GGTGAAGAAATGATGATTAATTTCTAAGATATGCGAAATGCTTTTATAGAATGACG 342  
 Db 272 GlyGlnIleGlyAspGlnValIleLeuArgIlyArgLeuThrSerTyrGlnAsnGlnLeu 291  
 QY 343 GATTTGAAGTCATATCTCTCAAAAACATTTGGAAATGT---CCTGAGCTTGAAT--- 396  
 Db 292 GlnIleAsp---GlnLeuGlnIleAspIleGlnThrCysAsnSerSerMetAlaAsnGln 310  
 QY 397 ---AAACCATGATGCTTATTTCAGATGACGATTAATTAATGACCTATATTTTCTCT 453  
 Db 311 ValGlnProIleSer-----LeuGlnLeuProPheSerSer 322  
 QY 454 GAA----- 456  
 Db 323 LeuThrAlaGlySerThrHisSerProGlnArgTyrGlnGlyMetLeuValIysLeuPro 342  
 QY 457 CAGAAATTCACATGAGAGAAATCAACAATCAATGCGCAAGATTTTAAATTAAT 516  
 Db 343 GlnThrLeuThrValSerGlnLeuTyr-----AsnTyrGlyArgTyrGlyGlnLeu--- 359  
 QY 517 GATTTCTTATTTACTTATCTGCTGTAACCTGACCTGGAGAGAGATTTTCAAAAAAC 576  
 Db 360 -----SerLeuSerLeuGlyArgLeuTyrIleProThrAsn 371  
 QY 577 TTTTCAATGATTA-----GAGGCTAATCATTGAGAAATTAATTAAGAGAAAAA 630  
 Db 372 LeuTyrProAlaLeuSerProGlnAlaIysAlaLeu-----AlaGlnIysAsnLeu 388  
 QY 631 CTTTCTAAACCTTTCTTT-----CGACCAACCGCAGAGATTA 666  
 Db 389 LeuSerLeuIleIlePheAspAspGlyTyrAsnAsnGlnAsnArgThrProTrp---Leu 407  
 QY 667 CCTGATGCG-----AGATAGATTATTTGCTGCGACCAACAGAA 705  
 Db 408 ProThrAsnPheSerValAlaAsnThrLeuArgSerGlyTyrGlnLeuIysAsnValGln 427  
 QY 706 GCG-----CCTAAATGAGAGATG--- 723  
 Db 428 GlyIleLeuGlnTyrArgPheAsnGlyTyrPArgValGlnProValLeuGlyArgThrGln 447  
 QY 723 ----- 723  
 Db 448 ProGlnValIleThrGlnThrAsnProAlaGlnAsnIleIleThrIysAsnAlaAsnHis 467  
 QY 724 -----AGTTTAAAGAACTTAAATAATCAACAATCTAGAATGATTTCTAAT 771  
 Db 468 IleArgValAlaSerPheAsnValIleuAsnTyrAspAsnGlyAlaThrGlyPheProThr 487  
 QY 772 ATGGAAGGGCTGCAAAACAA-----AAGTATAGTTCAATT 807  
 Db 488 GlnArgGlyAlaAsnThrGlnAlaGlnPheAspIysGlnHisIleIysIleValSerAla 507  
 QY 808 ATAAAAGAGGTA-----CAAAAGGTAACGCT 834  
 Db 508 LeuIysSerIleAspAlaAspValTyrGlyLeuMetGlnIleAlaAsnAsnGlyTyrGly 527  
 QY 835 CCACAGACAGACGAAAGATTTGGTACACCGCAGTGC----- 873  
 Db 528 ProAsnSerAlaIleAlaHisLeuThrSerAlaLeuGlyProAspTrpIysTyrValIle 547  
 QY 874 ---AGTAACCTGGAATAATTTCCGATTAATTTATATAGTGTAGG--- 915  
 Db 548 ProGlnAsnLeuAspArgLeuGlyAsnAspValIleAlaValAlaIleIleTyrAsnSer 567  
 QY 916 -----CTAAGCCAAAAGACAGAGTA 936  
 Db 568 LysArgValIysProLeuAsnIysAlaValValLeuAspLeuGlyAspIysAsnArgThr 587  
 QY 937 ACCTTACTCA----- 948  
 Db 588 ThrLeuAlaGlnThrPheGlnAlaValArgGlyAsnIysIlePheThrValIleProAsn 607

QY 949 -----AATGATGACATAACA 966  
 Db 608 HisLeuIysSerIysGlyCysSerGlyValAlaAspIleAsnSerSerSerAlaAspGln--A 627  
 QY 967 ATGACGGTTCAATAGTGTGGAAAC 989  
 Db 627 snAspGlyGlnGlyCysTrpAsn 634  
 RESULT 6  
 US-09-248-796A-23971  
 ; Sequence 23971, Application US/09248796A  
 ; Patent No. 6747137  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Keith Weinstein et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
 ; FILE REFERENCE: 107196.132  
 ; CURRENT APPLICATION NUMBER: US/09/248,796A  
 ; PRIOR FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: US 60/074,725  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: US 60/096,409  
 ; NUMBER OF SEQ ID NOS: 28208  
 ; SEQ ID NO 23971  
 ; LENGTH: 600  
 ; TYPE: PRT  
 ; ORGANISM: Candida albicans  
 ; FEATURE:  
 ; NAME/KEY: UNSURE  
 ; LOCATION: (590), (595), (597)  
 ; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno  
 US-09-248-796A-23971  
 Alignment Scores:  
 Pred. No.: 0.0435 Length: 600  
 Score: 101.50 Matches: 62  
 Percent Similarity: 39.56% Conservative: 46  
 Best Local Similarity: 22.71% Mismatches: 110  
 Query Match: 5.85% Indels: 55  
 DB: 4 Gaps: 13  
 US-10-617-962-2 (1-1008) x US-09-248-796A-23971 (1-600)  
 QY 202 CCTGATGACTATTATAG-----AATAAAGACCTGCTGAGACA 240  
 Db 286 ProPheSerPheTyrGlnSerAlaSerAlaAsnSerSerAsnSerSerLeuIysThrIys 305  
 QY 241 ATTATCAAGATATATATGCTAATCTTATCATCTGACATATTA-----GGTGAAGAT 291  
 Db 306 ThrIleGlnIysThrLeuSerAsnSerSerSerSerSerLeuPheAlaThrProArgSer 325  
 QY 292 GGTGATCAAAATTTCTAAGATATGCAAAATGCTTTTATTAAGATGAACTGGATTTTGA 351  
 Db 326 GlyAlaIysAlaSerArg-----TyrIysAsnIleTyrAsp----- 337  
 QY 352 GGTCAATATCTCAAAAACATTTGGATGTTCTCTGACCTTGAATAAATCAATGAGTGT 411  
 Db 338 -----GlnSerIysIleIleSerAsnAlaMetIysIleSerSerGlnAsnValSerGly 355  
 QY 412 TATTCAGATGACGATTAATTAATTAAGACATATTTTCTGTCAGAGAAATTCACATG 471  
 Db 356 IleGln-----ArgThrValThrProProLeu 364  
 QY 472 GAGAAATCAACAATCAAAATGCCGACAGATTTTAAATTAATGATTTCTTATTTACC 531  
 Db 365 AsnArgAsnAsnAsnAsnAsnSerAsn-----LysTyrLeuAsnHis 379  
 QY 532 TTATCTGCTGATCTTCACTG---GGAAGAGAGATTTTCAAAAACCTTTACATGGA 588  
 Db 380 SerSerSerLeuProSerIleuIysGlyIysArgSerPheSer----- 393

```

QY      589 TTAGAGGCTAAATCATTAAGAAATTATTTAGAGAGAAAAAATTCTTAACCT----- 642
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      394 ---TyrValArgGlnLeu---ArgTyrGlyAspSerLysArgValAlaSerProLysThr 411
QY      643 -----TTCTTTCGACCGCAGAGATTACCTGATGGCAGAAATAGATTATTTGGCTGGA 696
      DB      412 AenIleThrIleAspThrProPheThrProPheThrAlaThrAlaThrAlaThrAlaThr 431
QY      697 CCAGAGAGAGCGCTAAATGAGAGTGTGTTTAAAGAACTTAAATAACAATCTAG 756
      DB      432 AlaThrThrValProThr---SerValSerAlaValThrValLeuSerSerIleGlyLeu 450
QY      757 AATGATTTTCTAATATGAGAGGGCTGCAGAAAACAAGTATGCTCATTTATTAAGAG 816
      DB      451 SerGly-----SerProSerSerLysSerProGlnThrGlnPheIleLeuGln 466
QY      817 GTACAAAGAGGTAAAGCTCACAGCAGAGGAGAAAGTATGCTACGCGAGTGGCAGT 876
      DB      467 LeuProSerAlaSerThrProSerThrSerAlaThrThrLeuSerThrSerSerAlaSer 486
QY      877 AACCTGAGAAATTTGCCGATATTTATATAGTGTGAGGCTTAAGCCAAAAGACAGAGGTA 936
      DB      487 GlnLeuLysLysAsnArgAsnSerProValSerValAsnSerGlnAlaLeuAspSerTyr 506
QY      937 ACCTTACTCAAAATGATATCTGACAAATACATGACGCTT 975
      DB      507 Thr---ThrSerSerSerThrAsnSerThrMetAspIle 518

```

## RESULT 7

```

US-08-628-291-4
; Sequence 4, Application US/08628291
; Patent No. 5801031
; GENERAL INFORMATION:
; APPLICANT: GALIVAN, JOHN H.
; APPLICANT: RYAN, THOMAS J.
; APPLICANT: YAO, RONG
; APPLICANT: NIMEC, ZENIA
; TITLE OF INVENTION: GAMMA GLUTAMYL HYDROLASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,291
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: TIMIAN, SUSAN J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1636
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 293 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-628-291-4

```

## Alignment Scores:

```

Pred. No.: 0.113          Length: 293
Score: 96.50             Matches: 51
Percent Similarity: 40.57%  Conservative: 35
Best Local Similarity: 24.06%  Mismatches: 99

```

```

Query Match: 5.56%          Indels: 27
DB: 1          Gaps: 11
US-10-617-962-2 (1-1008) x US-08-628-291-4 (1-293)

```

```

QY      244 TTACAGAAATATATGTCTAATCTTCATCTGACATTAATAGCGAAATGCGATCAATT 303
      DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      58 TyrGlnThrLeuPheArgSerIleAsnGlyValLeuLeuProGlyGlyValAlaAsnLeu 77
QY      304 TCTAAAGAT-----ATGCGAAATGGTTTATTAAGATAAGTACGTGATTTT--- 348
      DB      78 ThrIleSerGlyTyrSerArgValAlaLysIlePhePheThrLysAlaLeuGlnSerPhe 97
QY      349 ---GAAGTCAATAT---CTGCAAAACATTGGAAATGTTCTCGACCTTGAATAAACCA 402
      DB      98 AspAsnGlyAspTyrPhePro-----ValTrpGlyThrCys---LeuGlyLeuGlnGlu 114
QY      403 TTGAGTGTCTTATTCAGATGACGATAAATTTATTAAGCATATATTTTCTCTGACAGAA 462
      DB      115 LeuSerValLeuValSerAsnAspAsnLeuThrLeuThrAsnThrSerSerValLys 134
QY      463 ATTCCACTGAGAGAAATCAACAATCAAAATGCGCAGAGATTTTAAATTAATGATTTC 522
      DB      135 LeuProLeuAsnPheThrArgAspSerLysGlnSerArgMetPheArgAsnLeu----- 152
QY      523 TTATTTACCTTATCTGTCTGTAACTTCACTGGAGAGAGATTTTCAAAAACTTTTAC 582
      DB      153 -----ProGlnGlnLeuLeuAsnSerLeuAlaSerGluAsnLeuThrAlaAsnPheHis 170
QY      583 AATGATTAAGAGCTTAATCATTAAGAAATTATTTAGAGAGAAAAAATTCTTAACCT 642
      DB      171 LysTrp-----SerLeuSerValLysAsnPheThrGlnAsnGlnLysLeuLys--- 187
QY      643 TTCTTTCGACACCGCAGAGATTACCTGATGGCAGAAATAGTTATTTGGCTGGA----- 696
      DB      188 PhePheAsnIleLeuThrValAsnThrAspGlyLysThrGlnPheIleSerSerMetGlu 207
QY      697 -----CCAGAGAGCGCTTAATGAGAGTGTGTTTAAAGAACTTAATAAT 744
      DB      208 GlyTyrLysTyrProIleTyrAlaValGlnTrpHisProGlnLysAlaProPheGlnTrp 227
QY      745 AACAAATCTAGAAATGATTTTCTAATATGAGAGGGCTGCAGAAACAAGAT----- 798
      DB      228 LysLysLeuArg---GlyLysSerHisAlaProAsnAlaValLysThrSerPheTyrLeu 246
QY      799 ---AGTTCAATTATTAAGAGGTACAAAAGGTAAC 831
      DB      247 AlaLysPhePheIleSerGlnAlaLeuLysAsnAsp 258

```

## RESULT 8

```

US-09-128-722-4
; Sequence 4, Application US/09128722
; Patent No. 5962235
; GENERAL INFORMATION:
; APPLICANT: Galivan, John H
; APPLICANT: Ryan, Thomas J
; APPLICANT: Yao, Rong
; APPLICANT: Nimec, Zenia
; TITLE OF INVENTION: Gamma Glutamyl Hydrolase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jaeckle Fleischmann & Muegel, LLP
; STREET: 39 Steele Street
; CITY: Rochester
; STATE: New York
; COUNTRY: US
; ZIP: 14614-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```



	APPLICATION NUMBER:	US/09/128-722
	FILING DATE:	04-AUG-1998
	CLASSIFICATION:	514
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	US 08/628,291
	FILING DATE:	05-APR-1996
	ATTORNEY/AGENT INFORMATION:	
	NAME:	Braman, Susan J
	REGISTRATION NUMBER:	34,103
	REFERENCE/DOCKET NUMBER:	87681.98R196
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE:	716-262-3640
	TELEFAX:	716-262-4133
	INFORMATION FOR SEQ ID NO:	4:
	SEQUENCE CHARACTERISTICS:	
	LENGTH:	293 amino acids
	TYPE:	amino acid
	STRANDEDNESS:	not relevant
	TOPOLOGY:	linear
	MOLECULE TYPE:	protein
	US-09-128-722-4	
	Alignment Scores:	
	Pred. No.:	0.113 Length: 293
	Score:	96.50 Matches: 51
	Percent Similarity:	40.57% Conservative: 35
	Best Local Similarity:	24.06% Mismatches: 99
	Query Match:	5.56% Indels: 27
	DB:	Gaps: 11
	US-10-617-962-2 (1-1008) x US-09-128-722-4 (1-293)	
QY	244 TATCAAGATATATGCTTAATCTTTCATCTGCATAATTAGTGAATAAGCGTACAAATT	303
	:	
Db	58 TyrgIutHrLeuPheargSerileanglyValleuLeuProclglyglalaenLeu	77
	:::: :::::	:::
QY	304 TCTAAGAAT-----ATGGCAATGGTTTATTAAGAATGACTCGATTT--	348
	:::     ::	::::
Db	78 ThrHisSerIyTySerArgValAlaLysilePhePheThrLySalaleuGIuSerPhe	97
	:    ::	:
QY	349 ---GAAGGTCAATAT---CCTCAAAAATTTGGAAATGTTCCTAGCTTGAAAAATAACA	402
	::	
Db	98 AspaangIyaSprTyPhePro-----ValTrpGlyThCyS---LeugIyeGIuGIu	114
	::	
QY	403 TTGAGTGCTTATTCAGATGACGATTAATATTATTAAGCATATATTTTTCTGTACAGAA	462
	::	
Db	115 LeuSerValleuValSerasnAspaenleuLeuThrLeuthraSnhrSerSerVallys	134
	::    ::	
QY	463 ATTCACACTGGAGAAAAATCAACAATCAAAATGCCGCAAGATTTTAAATTAAATTGATTTC	522
	::    ::  ::	::::
Db	135 LeuProLeuasnPhethrArGaSpSerIySGInSerArgmetPheaArgenLeu-----	152
	::    ::  ::	::::
QY	523 TTATTACTTATCTGCTGTACTTACCTGACGCGGAGAGAGATTTTTCAAAAAATTTC	582
	::	
Db	153 -----ProGIuGluleuLeuasnSerleuAlaSerGIuasnleuThrAlaasnPheHis	170
	:::	
QY	583 AATGATTAGAGGCTAAATCATTTATAGATTAATATTATTAAGAGAAAAAATTCTTAAACT	642
	:::     ::	
Db	171 LysTrp-----SerleuSerVallysaSnPheThcIuasnIuIySleuIySls---	187
	::	
QY	643 TTCCTTCGACCAACCGCAGAGATTACTCTGATGCGAGAAATAGTTATTTGGCTGGA----	696
	::	::::
Db	188 PhePheasnIleleuThrValasnThrAspGIyLylThcIuPheIleSerSerMetGlu	207
	::    ::	::::
QY	697 -----CCAACAAGGCGCTTAATAGAGAGAGTGAAGTTTAAAGAATTAAAT	744
Db	208 GlyTyTLySLyTyProIleTyTralavalGIuTrpHisProGIuIySalapropheGIuTrp	227
	::	
QY	745 AACCAATTCAGAAATGATTTCTTAAATATGAAGGCGTCGCAAAACAAGAT-----	798
Db	228 LySlvSleuArg----GlylleSerHisalaproAsnAlavalLyshTriserPheTyLeu	246
	::	
QY	799 ---AGTTCAITTAATAAAGAGGTACAAAGGTTAAC	831

```

Db      247  AlAlyPhePheIleSerGluIaLeuLysAsnSp 258
RESULT 9
US-08-628-291-12
? Sequence 12, Application US/08628291
? Patent No. 5801031
GENERAL INFORMATION:
APPLICANT: GALIVAN, JOHN H.
APPLICANT: RYAN, THOMAS J.
APPLICANT: YAO, KONG
APPLICANT: NIMEC, ZENIA
TITLE OF INVENTION: GAMMA GLUTAMYL HYDROLASE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/628,291
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: TIMIAN, SUSAN J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/150
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1636
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-628-291-12
Alignment Scores:
Pred. No.: 0.116 Length: 317
Score: 96.50 Matches: 51
Percent Similarity: 40.57% Conservative: 35
Best Local Similarity: 24.06% Mismatches: 99
Query Match: 5.56% Indels: 27
DB: 1 Gaps: 11
US-10-617-962-2 (1-1008) x US-08-628-291-12 (1-317)
QY      244  TATCAAGAAATATATGTCATCTTCATCTGCATCTATTAGGTGAAATGTCATCAATT 303
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db      82  TygIuThrlleupheargserIleasnGlyValleuLeuproGlyGlyAlaAsnLeu 101
QY      304  TCTAAAGAT-----ATGCCAAATGCTTTTATTAAGATGAACGCGATTT--- 348
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      102  ThrHisSerGlyTyriserArgValalaIysIlePhePheThrLysAlaLeuGluSerPhe 121
QY      349  ---GAAGTCATAT---CCTCAAAACATTGGAGATGTTCTCGAGCTTGAAATTAACA 402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      122  AspaAsnGlyAspGlyrPhePro-----ValTrpGlyThrCys---LeuGlyLeuGluGlu 138
QY      403  TTGAGTGCCTTATTACAGATCAATAAATTATTAGCACTATATTTTCTCTGTACAGAA 462
      ||||| ||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
Db      139  LeuSerValleuValSerAsnAspAsnLeuThrLeuThrAsnThrSerSerValIys 158
QY      463  ATTCCACTGGAGGAAATCAACAACATCAATATGCGCAAGATTTTAAATTATGATTTTC 522
      ||||| ||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
Db      159  LeuProLeuAsnPheThrArgAspSerIleGlyAsnSerArgMetPheArgAsnLeu----- 176

```

```

QY 523 TTATTTACCTTATCTGCTGTAACTTCACTGGAGAGAGATTTTTCAAAACTTTTAC 582
Db 177 -----ProGlugluLeuLeuLeuSerLeuAlaSerGluAsnLeuThrAlaAsnPheHis 194
QY 583 AATGATTAGAGCTAAATCTAGAAATTATATTGAGAAAAAACTTTCTAACT 642
Db 195 LysTrp-----SerLeuSerValLysAsnPheThrGluAsnGluLysLeuLysLys--- 211
QY 643 TTTCTTCAGCACCGCAGAGATTCTGATGGCAGATAGATTATGGCTGGA----- 696
Db 212 PhePheAsnIleLeuThrValAsnThrAspGlyLysThrGluPheIleSerSerMetGlu 231
QY 697 -----CCAAAGAGCGCCTTAATGAGAGTGAAGTTTAAAGAACTTAAAAAT 744
Db 232 GlyTyrLysTyrProIleTyrAlaValGlnTrpHisProGluLysAlaProPheGluTrp 251
QY 745 AACCAATCTAGAAATGATTTTCTATATGAGAGGGGCTGCAAAACAAAGTAT----- 798
Db 252 LysLysLeuArg--GlyIleSerHisAlaProAsnAlaValLysThrSerPheTyrLeu 270
QY 799 ---AGTTCATTTTATAAAGAGGTACAAAGGGTAAAC 831
Db 271 AlaLysPhePheIleSerGluAlaLeuLysAsnAsp 282

RESULT 10
US-09-128-722-12
; Sequence 12, Application US/09128722
; Patent No. 5962235
; GENERAL INFORMATION:
; APPLICANT: Galivan, John H
; APPLICANT: Ryan, Thomas J
; APPLICANT: Yao, Rong
; APPLICANT: Nimec, Zenia
; TITLE OF INVENTION: Gamma Glucamyl Hydrolase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Jaeckle Fleischmann & Wogel, LLP
; STREET: 39 State Street
; CITY: Rochester
; STATE: New York
; COUNTRY: US
; ZIP: 14614-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,722
; FILING DATE: 04-AUG-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/628,291
; FILING DATE: 05-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Braman, Susan J
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 87681,98R196
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-262-3640
; TELEFAX: 716-262-4133
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-128-722-12

Alignment Scores:
Pred. No.: 0.116 Length: 317
Score: 96.50 Matches: 51

```

```

Percent Similarity: 40.57% Conservative: 35
Best Local Similarity: 24.06% Mismatches: 99
Query Match: 5.56% Indels: 27
DB: 2 Gaps: 11

US-10-617-962-2 (1-1008) x US-09-128-722-12 (1-317)

QY 244 TATCAAGATATATATCTTAATCTTCACTGTGCACTATTAAGTGAATAATGATCAAT 303
Db 82 TyrGluThrLeuPheArgSerIleAsnGlyValLeuLeuProGluGlyIleAlaAsnLeu 101
QY 304 TCTAAAGAT-----ATGCCAAATGCTTTTATTAAGAAATGAATCGATTTT--- 348
Db 102 ThrHisSerGlyTyrSerArgValAlaLysIlePhePheThrLysAlaLeuLeuSerPhe 121
QY 349 ---GAAGGCAATAT--CCTCAAAACATTTGGAATGTTCTCGACCTTGAATAAACA 402
Db 122 AspaAsnGlyAspTyrPhePro-----ValTrpGlyThrCys--LeuGlyLeuGlu 138
QY 403 TTGAGTGTCTTATTCAGATGACGATTAATTAATTAATTAATTTTCTGTACAGAA 462
Db 139 LeuSerValLeuValSerAsnAspAsnLeuLeuThrLeuThrAsnThrSerSerValLys 158
QY 463 ATTCACCTGAGAGAAATCAACATCAATCAATGCGCAAGATTTTAAATTAATGATTTC 522
Db 159 LeuProLeuAsnPheThrArgAspSerLysGlnSerArgMetPheArgAsnLeu----- 176
QY 523 TTATTTACCTTATCTGCTGTAACTTCACTGGAGAGAGATTTTTCAAAACTTTTAC 582
Db 177 -----ProGlugluLeuLeuAsnSerLeuAlaSerGluAsnLeuThrAlaAsnPheHis 194
QY 583 AATGATTAGAGCTAAATCTATTAAGATTAATTAATGAGAAAAAACTTTCTAACT 642
Db 195 LysTrp-----SerLeuSerValLysAsnPheThrGluAsnGluLysLeuLysLys--- 211
QY 643 TTTCTTCAGCACCGCAGAGATTACTGTGACAAATAGCTTATTTGGCTGA----- 696
Db 212 PhePheAsnIleLeuThrValAsnThrAspGlyLysThrGluPheIleSerSerMetGlu 231
QY 697 -----CCAAAGAGCGCCTTAATGAGAGTGAAGTTTAAAGAACTTAAAAAT 744
Db 232 GlyTyrLysTyrProIleTyrAlaValGlnTrpHisProGluLysAlaProPheGluTrp 251
QY 745 AACCAATCTAGAAATGATTTTCTATATGAGAGGGGCTGCAAAACAAAGTAT----- 798
Db 252 LysLysLeuArg--GlyIleSerHisAlaProAsnAlaValLysThrSerPheTyrLeu 270
QY 799 ---AGTTCATTTTATAAAGAGGTACAAAGGGTAAAC 831
Db 271 AlaLysPhePheIleSerGluAlaLeuLysAsnAsp 282

RESULT 11
US-09-538-092-119
; Sequence 119, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 119
; LENGTH: 623
; TYPE: PRP
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:

```

NAME/KEY: misc feature  
LOCATION: (0) (0)  
OTHER INFORMATION: Polypeptide Accession Number YDL203C  
US-09-538-092-119

Alignment Scores:  
Pred. No.: 0.167 Length: 623  
Score: 96.00 Matches: 74  
Percent Similarity: 36.93% Conservative: 56  
Best Local Similarity: 21.02% Mismatches: 114  
Query Match: 5.53% Indels: 108  
Gaps: 16

US-10-617-962-2 (1-1008) x US-09-538-092-119 (1-623)

```

QY 79 ATACTAACTTTAAGCAACAGATGAG-----GCTCATACACATCATAT 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 ILePheAsnSerLyProThrGlyLulIleGlyTyrAlaAsnAsnGlyThrAsnSerThr 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 GGAATTGAATATCGA-----GCTAAGAAATAATATTACTTAC 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 GlySerArgTyrGluLeuProPheAsnPheSerThrLySGluSerLeuGlySerPro 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 163 GCTTGGCTGTAAAGTGTATTCATATGATCTTAACCTTCGATGACATATTATTAAGAT 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 AlaValGlnAspAlaSerIleSerSerGlyAsnArgIleSerGluSerValArgAspAsn 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 223 -----AAGAGACTGCTGAGAGAAATTATCAAGATAT----- 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 SerAlaProProTyrGluGluSerGluSerArgIleLeuGlnGluValTyrArg 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 256 -----ATGCTCAT----- 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 ThrGluGluLysAlaProIleArgProLeuAsnAsnProValProProGlnLysIle 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 265 -----CTTTCATCTGCACTATTAGGTGAATAATGTCATCAATTTCTAAAGT 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 AsnGlnProProThrGlySerIleAlaLysThrAspAsnAsnGlySerSerGlyGlyGluAsp 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 313 ATGCGAAATGGTTTATTAAGATGAATGATTTTGAAGTCATATCTCAAAACATT 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 LysLeuSerSerTyrSerProGluAlaLeuAlaPheTyrGlnValTyrLysLeuThrIle 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 TGGAT-----GTTCTGAGCTGGAAT-----AAACCATTTGAGTGT 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 ThrAspSerSerLysPheThrProGluIleGlnMetGlnTyrCysGluThrLeuThr 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 412 TATTCAGATGACGATTAATTTATTAACACTATATTTTCTCTGACAGGAATT----- 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 TyrAlaPheAsnGluAspPheIleSerGlnTyrAsnIleAsnAlaGluLysLeuLysArg 287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 -----CCACTGAG-----GAAATCAACAATCAATGCGCAAGATTTTAA 510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 SerLeuLysProGluGlnMetLeuLysAsnGlnLysValIleLeuGlnLysSerPhe 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 511 TTAATGATTTTATTAATCTTATCT-----GCTGTAACTTCACTGGAGG--- 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 ValLeuThrLysLeuIleThrLeuLysTyrProProAlaMetLysLeuMetGlyThrLeu 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 559 -----AGATTTTTCAAAAAACTTTTAAATGATGATTAAGGCT 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 TyrSerHisGlnProTyrLeuProIleLysAsnLysAsnIleValIleLysAsnAspGlu 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 598 AAATCATTAAGAAATTAATTAAGAGAAAAAACTTTCTTAACTTTTTCGACACCG 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 LysAlaLeuGluTyrTyrCysLysAlaAlaLysLeuAsn----- 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 658 CAGAGATTACCTGATGCGAGATAGTTTGGCTGCA----- 696
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 -----AsnSerAspAlaCysTyrArgAlaGlyValCysPheGluTyrGlnArg 376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 697 -----CCAAAGAGCGCCTTAATGAGAGATGATTTTAA 732
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 377 GlyThrSerSerLeuAspProSerProThrLysGluGlnCysIleLysLysAlaPheGln 396
QY 733 -----GAACCTTAATAATAACAATCT-----AGGAATGATTT 765
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 TyrTyrGlnHisGlyAlaGluValCysSerAsnSerAlaCysMetTyrLysLeuGlyMet 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 766 TCTAATATGAAAGGGCTGCAAAACAAG-----TATAGTTCATTT 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 SerHisLeuTyrGlyLysLeuAsnMetGlnTyrThrAspValLeuLeuAlaIleLysTrpPhe 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 808 ATAAAGAGGTACAAAGGGTACCGCTCCACAGACA 843
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 437 AspLysAlaAlaGlnLysGlyLysPheProGlnThr 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

# RESULT 12

```

US-08-951-871-2
; Sequence 2, Application US/08951871
; Patent No. 5866398
; GENERAL INFORMATION:
; APPLICANT: XU, SHUANG-YONG
; TITLE OF INVENTION: METHOD FOR CLONING
; TITLE OF INVENTION: AND PRODUCING THE BslI RESTRICTION ENDONUCLEASE IN E.
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: New England Biolabs, Inc.
; STREET: 32 Tozer Road
; CITY: Beverly
; STATE: MA
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,871
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Gregory D
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 978-927-5054
; TELEFAX: 978-927-1705
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; TOPOLOGY: linear
; FRAGMENT TYPE: internal
US-08-951-871-2
Alignment Scores:
Pred. No.: 0.192 Length: 912
Score: 96.00 Matches: 66
Percent Similarity: 34.84% Conservative: 42
Best Local Similarity: 21.29% Mismatches: 110
Query Match: 5.53% Indels: 92
Gaps: 12
US-10-617-962-2 (1-1008) x US-08-951-871-2 (1-912)
QY 4 GTTATACATTAACACCTGATGATAGAGTATTCACCGCTTGAAGCAATATGACA 63

```

```
Db      21 ValValSerLeuIleAspGluAsp-----AlaLeuSerLeuGlu 33
Qy      64 GGAGATATATAGTACTATACTTAAGCAAAACAGATGAGGTCATACGATCATAT 123
Db      34 GluGlnPheProLeuValLeuLeuSerProValValAspGlu----- 47
Qy      124 GGAATTAATATGAGCTTAAGAAATATATATAGCTTACGCTTGGCTGAAGTGATAT 183
Db      48 -----GluIleValIleTyrIleLeuCysGluTyrIleAsnLeuAsnAlaLeuAsnVal 65
Qy      164 CATATGATCTTAACCTTCGATCATCTTATTAAG-----ATAAAGAGACTGCT 234
Db      66 LysThrIleSerGlnThrLeuAsnLysGluTyrLysPheGlyArgAsnSerLysThrAla 85
Qy      235 GAGAAATTTAT-----CAAGATATATGCTTAATCTTCATCGTCGACCA 279
Db      86 LeuLysLysTyrLeuAspTyrGlyLysGluGluTyrLeuIleGlnPhePheAsnThrLeu 105
Qy      280 TTAGGTGAATAATGTGATCAAAATTTCTAAAGATATGGCAAAATGCTTTTATTAAGATGA 339
Db      106 MetLeuGluAsnAsnThrTyrIleAspArgGlu-----TyrIleGluSerVal 121
Qy      340 CTGGAATTTGAAGTCATATCTCAAAACATTGGAAATGCTTCGAGCTTGAATAA 399
Db      122 LeuAlaPhe-----CysGlu 126
Qy      400 CCATTGAGCTTAATTCAGATGACGATAAATTATTAACATATATTTTCTCTGACAG 459
Db      127 ProValSerLysGluLysIleLysAsnGluPheIleLysLeuTyr----- 141
Qy      460 GAAATTTCCAGTGAAGAAATCAACATCAAAATGCCGAAAGTTTATTAATTAATGAT 519
Db      142 -----AsnGluAlaAsnGluValAsnGluTyrGlyLysLeuLysAsp 155
Qy      520 TTCTTATTTACTTATCTGCTGTACTTCACTGGGAAGAGATTTTCAAAAACTTCT 579
Db      156 TyrLeu-----LeuGly-----IleTyrSerLysLeuPhe 165
Qy      580 TACAATGATTAAGAG---GCTAAATCATTAAGAAATTAATTAAGAAAAAACTTCT 636
Db      166 SerMetGlyLeuGlnAsnLeuArgLeuIleGluIleTyrAsnSerAsnGluSerLeuIle 185
Qy      637 AAACCTTTCTTCGACCAACCGAGATTACTGATGCGAATAGATTATTTGGCTGA 696
Db      186 LysLysValPhe----- 189
Qy      697 CCAACGAAAGCCCTAAATGAGAGTGAAGTTTAAAGACTTAAATAACAATATAG 756
Db      190 -----LysTyrGluSerThrIleLysGluLeuLysGluTyrCysLeuSer 204
Qy      757 AATGATTTTCTAATATGAGAGGGGCTGCMAAACAAAGATATAGTATTAATAAAGAG 816
Db      205 AsnGlnGluSerIleThrAlaGlyLeuAlaIleLysMetPheAsnGluLysTyrMetGlu 224
Qy      817 GTACAAAAGGATGAGCTTCACAGACAGACGAGCGAAAGTATGTTACAGCCAGTGCACT 876
Db      225 LeuMetLysLysGluTyrGlnGlnAspAlaIle-----AlaLeu 237
Qy      877 AACCTGMAAAATTCGCCGAATTAATTAAT 906
Db      238 LysLeuGlnGluIleMetAsnGlnLeuTyr 247

RESULT 13
US-09-489-039A-14067
; Sequence 14067, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
```

```
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14067
; LENGTH: 3168
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14067

Alignment Scores:
Score: 0.301 Length: 3168
Percent Similarity: 96.00 Matches: 74
Best Local Similarity: 33.42% Conservative: 49
Query Match: 20.11% Mismatches: 123
Db: 5.53% Indels: 122
Gaps: 18

US-10-617-962-2 (1-1008) x US-09-489-039A-14067 (1-3168)
Qy      76 CGTATCTTAAC-----TTAAGCAAAACAGATGAGGTCATACA 114
Db      981 ArgLeuLeuAsnAsnCysValValAspGlyAspTyrArgCysThrAspGlyArgTyrVal 1000
Qy      115 GCATCATATGGAATTTGAATATGACCTAAGAAATAATATTA-----GCTTAC 162
Db      1001 ArgAlaArgProIleGlnIleAsnGlnIleArgGluSerLeuThrGluLeuAlaGlyTyr 1020
Qy      163 GCTTTGGCTGAAGGATTT---CATATGATCTTAACCTCGATGACATATATAG 219
Db      1021 CysGluGlyPheGlnAlaIleProAspThrIleAlaArgAlaGlyAspArgLeuTyrGlu 1040
Qy      220 AATAAGAGACTGCTGAGAGA----- 240
Db      1041 MetMetSerGlyAlaGluGluProValAlaIleIlePheProGlnSerAlaSerAspGly 1060
Qy      241 -----ATTATCAAGATATATATGCTAATCTTCACTGCACTATTAAGTGAAT 291
Db      1061 ValGluValLeuTyrGlnGluPheSer-----PheGlyArgTyr 1073
Qy      292 GGTGATCAAAATTTCAAGATATGCGAAATGTTTAAAGATGAATGATTTTGA 351
Db      1074 PheAsnGlnIleAlaIleGlyValLeuArgGlyIleValGlnThrArg-----Gln 1090
Qy      352 GGTCAATATCTCAAAACATT-----TGG 375
Db      1091 ProArgGlnProLeuArgIleLeuGluValGlyGlyLeuThrGlyTyrThrAlaTyr 1110
Qy      376 AATGTTCTGAGCTTGAATAATTAACCATTAAGTGTATTTCAGATGACGATTAATTA 435
Db      1111 LeuLeuProGluLeuAsnGlyValProAlaLeuGluTyrHisPheThrAsp---IleSer 1129
Qy      436 GCATCATAT-----TTTTCTCTGTACAG 459
Db      1130 AlaLeuPheThrArgArgAlaGlnGlnLysPheAlaAspTyrAspPheValIleTyrSer 1149
Qy      460 GAAATTTCACTGAGAGAAATCAACATCAATCAATGCC-----GCAAGATTTTAAATTA 513
Db      1150 GluLeuAspLeuGluLysGluAlaGlnSerGlnGlyPheGlnAlaGlnSerTyrAspLeu 1169
Qy      514 ATGATTTCTTATTAACCTTAATCTGCTGTAACTTCACTGGAAGAGAGATTTTCAAAA 573
Db      1170 IleValAlaAlaAsnValIleHisIleThrArgHisIleGlyArg----- 1184
Qy      574 AACTTTACAATGATTAAGAGCTTAATCACTTAAGATTAATTAATGAGAAAAAAACTT 633
Db      1185 -----ThrLeuAspAsnLeu----- 1189
Qy      634 TCTAACTTTCTTCGACCAACCGAGAGATTA-----CCTGAT 672
Db      1190 ---ArgProLeuLeuLysProGlyLysArgLeuLeuMetArgGluIleThrGlnProMet 1208
Qy      673 GGCAGATATAGTTATTTGGCTGAGCAACAGAAAGCCCTTAATTAAGAGTGAATTTTAA 732
```

```

Db      1209 ArgLeuPheAspPheValPheGlyProLeuValLeuPro-----1221
Qy      733 GAACCTTAATAAATACAAATCTTAGAATGA-----TTTCTAATATGAGAGGCT 783
Db      1222 ---LeuGlnAspLeuAspAlaArgGluGlnGluLeuPheLeuThrAlaGlnTrpGln 1240
Qy      784 GCAAAACAAAGATAGTTCATTATTAAGAGATACAAAGAGATACGCTCCACAGACA 843
Db      1241 GlnGlnCysArgHisAlaGlyPheSerIysValAlaTrp-----LeuProGlnAsp 1257
Qy      844 GCAGCAAAAGATATTGTATGACAGCGACGACGATACCTGCAAAAATTGCCAATATTTA 903
Db      1258 GlySerProThrAlaGlyMetSerGluHisIleIleLeuAlaThrLeuProGlyGlnAla 1277
Qy      904 TATAGTGTAGGCTAAGCCAAAAGACAGGGTAACTTTACT-----945
Db      1278 ValSerAla-----ValThrPheThrAlaProSerGluProVal 1290
Qy      946 -----CAAAATGATACCTGACAT 963
Db      1291 LeuGlyGlnAlaLeuThrAspAsn 1298

RESULT 14
US-09-583-110-3191
; Sequence 3191, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3191
; LENGTH: 680
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3191

Alignment Scores:
Pred. No.: 0.195      Length: 680
Score: 95.50      Matches: 81
Percent Similarity: 39.89%      Conservative: 65
Best Local Similarity: 22.13%      Mismatches: 123
Query Match: 5.50%      Indels: 97
DB: 4      Gaps: 22

US-10-617-962-2 (1-1008) x US-09-583-110-3191 (1-680)
Qy      25 GATAGAGTGAATATCCACCGCTTGAAAAGCAATAGACGAGATATAGTACGTACTA 84
Db      65 AapAlaSerGlyLysProLeuValGluAenThrLeuLys-----GlnValVal 80
Qy      85 AACCTTAAGCAAAACAGATAGGCTCATACGATCATATGAGTAATGATAGAGTAAAG 144
Db      81 SerPheThrArgSerAsnLys-----MetThrAlaThrAspLeuLysGluThrAlaLys 98
Qy      145 AAAATATATATAGCTTACGCTTGTAGGCTGATTCATCATATATATGCTTAACTTCT 204
Db      99 Lys---LeuLeuThrTyrlValSerIleSerSerProAsnLeuThrGlnArgGlnLeuAla 117
Qy      205 GATGACTATATTAAGATAAAGAGACTGCTGAGAGAAATTTATCAGAAATATATGCTTAA 264
Db      118 AspTyrlTyrlLeuAlaAspProGlu-----IleTyrlLysThrValGlnAla 133

```

```

Qy      265 CTT---TCATCTGCATATTAGTGAAAGATGGTATCAATTTCTTAAGATATGGCAAT 321
Db      134 LeuProSerGluLysArgLeuAspSerAspIlysnAlaArgLeuSerGlu-----Ser 150
Qy      322 GCTTTTATTAAGATGAACCTGATTTTGAAGTCAATATCTCAAAACATTTGGAATGTT 381
Db      151 GluLeuTyrlAsnAlaValAlaAsp-----158
Qy      382 COTGAGCTTGAATAATTAACATAGTCTTATTCAGATGACGATTAATTAATGACGTA 441
Db      159 ---SerValGlnThrSerGlnLeuAsn---TyrlThrGluAspGluLys---LysGluIle 175
Qy      442 TATTTTCTCT-----GTACAGAAAT 465
Db      176 TyrlLeuPheSerGlnLeuAsnAlaValAlaGlnPheAlaThrGlyThrIleAlaThrAsp 195
Qy      466 CCACTGGAGAAATCAACATCAATCCGACAGATTTTAAATTAATGATTTCTTA 525
Db      196 ProLeuAsnAspSerGlnValAlaValIleAlaSerIleSerLysGluMetPro-----213
Qy      526 TTTACCTTATCTGCTGATACCTTCACTGAGAGAGATTTTCAAAAACCTTTACAT 585
Db      214 ---GlyIleSerIleSerThrSerTrpAspArgLysValLeuGluThrSerLeuSerSer 232
Qy      586 -----GATTAAGGCTTAATCATTAAGATTAAT 615
Db      233 IleValGlySerValSerSerGluLysAlaGlyLeuProIleGluAlaGluAlaTyrl 252
Qy      616 ATTGAGAGAAAAAACTTTCTTAACCTTTCTTTCGACACCGCAGAGATTAACCTGATGC 675
Db      253 LeuLysLys-----GlyTyrlSerLeuAsnAsp---261
Qy      676 AGAATAGT-----TATTTGGCTGACCA---ACAGAAAGCCCTTAATGAGAGTGA 726
Db      262 ArgValGlyThrSerTyrlLeuGluLysGlnTyrlGluGluThrLeuGlnGlyLysArgSer 281
Qy      727 TTTAAGAACTTAATAATTAACATCTAGAAATGATTTTCTAATATGAGAGGCTGCA 786
Db      282 ValLysGluIleHisLeuAspLys-----TyrlGlyAsnMetGluSerVal---296
Qy      787 AAACAAAGATATAGTTCATTATTAAGAGGTACAAAAGGTAAC-----831
Db      297 -----AspThrIleGluGlnGlySerLysGlyAsnAsnIleLysLeuThr 311
Qy      832 -----GCTCCACAGACGACGCAAAAGT---ATTGTAACGCCAGCTGCGATAC 879
Db      312 IleAspLeuAlaPheGlnAspSerValAspAlaLeuLysSerTyrlPheAsnSerGlu 331
Qy      880 CTGAA-----AAATTGCCGATATATTTATATAGTGTAGGCTTAAGCCAAAAA 927
Db      332 LeuGluAsnGlyGlyAlaLysTyrlSerGluGlyValTyrlAlaValAlaLeuAsnProLys 351
Qy      928 GACAGGTAACCTTACTCAAAATGATACGACAT-----ACAAATGACG 972
Db      352 ThrGlyAlaValLeuSerMetSerGlyIleLysHisAspLeuLysThrGlyGluLeuThr 371
Qy      973 GTTCAATGTTGGAAT 990
Db      372 ProAspSerLeuGlyThr 377

RESULT 15
US-08-328-322-5
; Sequence 5, Application US/08328322
; Patent No. 5723436
; GENERAL INFORMATION:
; APPLICANT: Huang, Laiqiang
; APPLICANT: Cyert, Martha S.
; TITLE OF INVENTION: Calcineurin Interacting Protein Compositions
; TITLE OF INVENTION: and Methods
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250

```

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,322  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: P38,615  
REFERENCE/DOCKET NUMBER: 8600-0151.10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-328-322-5

Alignment Scores:  
Pred. No.: 0.217 Length: 915  
Score: 95.50 Matches: 57  
Percent Similarity: 39.21% Conservative: 32  
Best Local Similarity: 25.11% Mismatches: 69  
Query Match: 5.50% Indels: 69  
DB: 1 Gaps: 14

US-10-617-962-2 (1-1008) x US-08-328-322-5 (1-915)

QY 34 GGATATCCACCCGTTGAAAGCAATAGCAGAGATATAGTACGTAACTAAAC----- 87  
||| ||||| : : : : : ||||| : : : : :  
Db 97 GlyIleProProluArgIluGluPheSer---AspValGluThrValAlaSerHisThr 115  
88 -----TTTAAACAACAGATGAGGCTCATACGATCATATGAAATGAAATGAGCT 141  
||| ||||| : : : : : ||||| : : : : :  
Db 116 TrpProPheTyrGlnAlaAspAspGlyMetAsn---SerPheThrLeuGluHisSer 134  
142 AAGAAATATATATTAGCTTACGCTTGGCTGAAGTGATTCATTAATGATCTAACTT 201  
135 Ser-----AsnAsnSerSerAsnArgProSerMet 144  
202 CCTGATGACTATTATTAAGATTAAGAGAGAGCTGAGAGATTAT----- 246  
||| : : : : : : : : : : : : : : :  
Db 145 SerAspGluAspTyrLeuLeuGluIuYserGlyAlaSerValTyrIleProThrAla 164  
247 -----CAAGAAATATATGCTTAATCTTTCATCTGCACTATGAGTAAATGGTATCAA 300  
: : : : : ||||| : : : : :  
Db 165 GluProProluAspAsnSerAsnLeuSer----- 174  
301 ATTTCTAAAGATATGCAATATGTTTATTAAGATGAATGAGATTGGAAGTCAATAT 360  
175 -----LeuAsp-----AlaTyr 178  
361 CCTCAAAACATTGGATGTCTCTGAGCTTGAATAAATCAATGAGTCT-----TAT 414  
: : : : : : : : : : : : : : :  
Db 179 GluArgAsnSerLeuSerSerAspAsnLeuSerAsnLysProValSerSerAspValSer 198  
415 TCAGATGACGATTAATTTATAGACTATATTTTCTCTGACAGAAATTCACCTGAG 474  
||| ||||| : : : : : ||||| : : : : :  
Db 199 HisAspAspSerLysLeuLeuAla-----IleGlnLysThrProLeuPro 213  
475 GAAATCAACAA-----TCAATGCCGCAAGATTTTAAATTAATTAATTTCTTATTT 528  
: : : : : : : : : : : : : : :  
Db 214 SerSerSerArgArgGlySerValProAlaAsnHisGly-----Asn 228

QY 529 ACCTTATCTGCTGTAACCTTCACTGGGAAGAGATTTTTCAAAAAATTTTACATGCA 588  
: : : ||||| : : : : : : : : : : :  
Db 229 SerLeuSerProHisThrPheIleSer---AspLeuPheThrIleThrPheSerAsnSer 247  
QY 589 TTAGAGGCTAAATCATTTAGAG-----AATTATATTGAGAGAAAAAATTTCTTAA 639  
||| ||||| : : : : : ||||| : : : : :  
Db 248 GlyAlaThrProSerProGluGlnGluAspAsnTyrIleuThrProSerLysAspSerLys 267  
QY 640 CCT-----TTCTTTCGACCA 654  
||| ||||| : : : : :  
Db 268 GluValPheIlePheArgPro 274

Search completed: November 21, 2004, 08:54:59  
Job time : 40.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 21, 2004, 08:44:30 ; Search time 89 Seconds  
(without alignments)  
8021.587 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1735  
Sequence: 1 atggttatacaataaacacc.....ctcattataaaatatataga 1008

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 3141230

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool.p/US10617962/runcat.17112004.163923.9785/app.query.fasta.1.1159  
-DB=Published Applications AA -QMT=faetan -SUFFIX=rapb -MIMATCH=0.1  
-LOOPT=0 -LOOPTXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MITLEN=0  
-MAXLEN=200000000 -USERS=US10617962.@CEN.1.1.130.@runcat.17112004.163923.9785  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
1	1734	99.9	335	15	US-10-617-962-4	Sequence 4, Appli
2	1276	73.5	368	15	US-10-617-962-3	Sequence 3, Appli
3	110	6.3	652	17	US-10-437-963-149602	Sequence 149602,
4	102.5	5.9	1581	16	US-10-425-115-293701	Sequence 293701,
5	101.5	5.9	1639	14	US-10-087-464-10	Sequence 10, Appli
6	100.5	5.8	1921	17	US-10-425-115-293718	Sequence 293718,
7	100	5.8	297	17	US-10-425-115-545697	Sequence 344697,
8	98.5	5.7	367	14	US-10-369-493-19786	Sequence 19786, A
9	98.5	5.7	616	15	US-10-282-122A-76864	Sequence 76864, A
10	98.5	5.7	1162	16	US-10-437-963-200517	Sequence 200517,
11	98	5.6	954	15	US-10-282-122A-54429	Sequence 54429, A
12	98	5.6	954	15	US-10-282-122A-63608	Sequence 63608, A
13	97.5	5.6	1770	17	US-10-425-115-271828	Sequence 271828,
14	97	5.6	878	17	US-10-425-115-249308	Sequence 249308,
15	97	5.6	1477	16	US-10-437-963-115886	Sequence 115886,
16	96.5	5.6	685	9	US-09-815-242-13477	Sequence 13477, A
17	96.5	5.6	685	15	US-10-282-122A-74118	Sequence 74118, A
18	96.5	5.6	685	16	US-10-474-776-710	Sequence 710, App
19	96	5.5	879	15	US-10-282-122A-54288	Sequence 54288, A
20	96	5.5	1006	16	US-10-437-963-192563	Sequence 192563,
21	95.5	5.5	303	15	US-10-282-122A-51836	Sequence 51836, A
22	95.5	5.5	360	15	US-10-282-122A-54381	Sequence 54381, A
23	95.5	5.5	414	14	US-10-369-493-22593	Sequence 22593, A
24	95.5	5.5	653	16	US-10-437-963-186821	Sequence 186821,
25	94.5	5.4	507	15	US-10-425-114-57458	Sequence 57458, A
26	94.5	5.4	611	15	US-10-425-114-54378	Sequence 54378, A
27	94.5	5.4	650	17	US-10-425-115-353793	Sequence 353793,
28	94.5	5.4	696	9	US-09-817-647-23	Sequence 23, Appli
29	94.5	5.4	696	9	US-09-877-665-23	Sequence 23, Appli
30	94.5	5.4	696	13	US-10-136-573A-23	Sequence 23, Appli
31	94.5	5.4	696	14	US-10-215-862-23	Sequence 23, Appli
32	94.5	5.4	1692	17	US-10-425-115-214669	Sequence 214669,
33	94	5.4	386	15	US-10-282-122A-54544	Sequence 54544, A
34	94	5.4	534	9	US-09-801-366-124	Sequence 124, App
35	94	5.4	737	16	US-10-437-963-107270	Sequence 107270,
36	94	5.4	910	15	US-10-456-129-7	Sequence 7, Appli
37	94	5.4	1009	15	US-10-607-631-2	Sequence 2, Appli
38	94	5.4	1097	14	US-10-369-493-1577	Sequence 1577, Ap
39	93.5	5.4	344	15	US-09-820-843A-35	Sequence 35, Appli
40	93.5	5.4	411	15	US-10-282-122A-53440	Sequence 53440, A
41	93	5.4	924	14	US-10-369-493-18383	Sequence 18383, A
42	93	5.4	1830	15	US-10-424-599-147146	Sequence 147146,
43	93	5.4	2872	17	US-10-783-528-116	Sequence 116, App
44	93	5.4	6713	15	US-10-282-122A-43811	Sequence 43811, A
45	92.5	5.3	539	16	US-10-408-765A-1437	Sequence 1437, Ap

#### ALIGNMENTS

RESULT 1  
US-10-617-962-4  
; Sequence 4, Application US/10617962  
; Publication No. US2004005036A1  
GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
; APPLICANT: EAST, Peter David  
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photo:  
; TITLE OF INVENTION: Luminescens  
; FILE REFERENCE: 050179-0076  
; CURRENT APPLICATION NUMBER: US/10/617,962  
; CURRENT FILING DATE: 2003-07-14  
; PRIOR APPLICATION NUMBER: US/09/463,048A  
; PRIOR FILING DATE: 2002-12-13  
; PRIOR APPLICATION NUMBER: PCT/AU98/00562  
; PRIOR FILING DATE: 1998-07-17  
; PRIOR APPLICATION NUMBER: PO 8088  
; PRIOR FILING DATE: 1997-07-17  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 335

```
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-617-962-4

Alignment Scores:
  Pred. No.:      3,116-161      Length:      335
  Score:          1734.00        Matches:     335
  Percent Similarity: 100.00%    Conservative: 0
  Best Local Similarity: 100.00%  Mismatches:  0
  Query Match:     99.94%       Indels:      0
  DB:              15           Gaps:         0

US-10-617-962-2 (1-1008) x US-10-617-962-4 (1-335)

QY 1 ATGGTATTAACAATTAAACCTGATGATAGATGATATATCAACCCCTGTAAGAAACAATA 60
DB 1 MetValIleGlnLeuThrProaspAspArgSerGlyTyrProProValGlnLysGlnIle 20
QY 61 GCAGAGATATGTCGTCGTCCTTAAACCTTAAAGCAAAACAGATGAGGCTCATACAGATCA 120
DB 21 AlaGlyAspIleValArgIleLeuAsnPhelySGlnThrAspGlnGlyHisThrAlaSer 40
QY 121 TATGGAATTTGAATTCGAGCTAAGAAATAATATTAGCTTACGCTTTGGCTGTAAGTGT 180
DB 41 TyrGlyIleGlnTyrArgAlaLysLysIleIleLeuAlaTyrAlaLeuAlaValSerGly 60
QY 181 ATTCAATTAATGATCTAAACCTTCCTGATGACTATTATTAAGAAATAAGAGACTGAGAGA 240
DB 61 ILeHtAsnValSerLysLeuProaspAspTyrTyrLysAsnLysGlnThrAlaGlnArg 80
QY 241 ATTATCAAGAAATATATGCTTAATCTTTGATCTGCATTTTACGTAAGAAATGGTGATCA 300
DB 81 IleTyrGlnGlnTyrMetSerAsnLeuSerSerAlaLeuLeuGlyGlnLysGlnLysGln 100
QY 301 ATTTCAAAAGATATGCGCAAAATGTTTATTAAGAAATGAACGATTTTGAAGTCAATAT 360
DB 101 IleSerLysAspMetAlaAsnGlnLysPheTyrLysAsnGlnLysAspPheGlnGlnTyr 120
QY 361 CCTCAAAACATTTGGAATGTTCCCTGAGCTTGAATAATTAACATTTAGTCTTATTCAGAT 420
DB 121 ProGlnAsnIleTyrAsnValProGlnLysGlnLysLeuLysProLeuSerAlaTyrSerAsp 140
QY 421 GACGATTAATTTTACACTATATTTTCTCTGTCGACGAAATTCACCTGAGGAAAT 480
DB 141 AspAspLysLeuLeuAlaLeuTyrPhePheSerValGlnGlnIleProLeuGlnGlnLys 160
QY 481 CAACAATCAATGCGCAAGATTTTATTAATTAATGATTTCTTATTAACCTTATCTGTGT 540
DB 161 GlnGlnSerAsnAlaAlaArgPhePheLysLeuIleAspPheLeuPheThrLeuSerAla 180
QY 541 GTAACCTTCACTGGAAGAGAGATTTTTCATAAAACCTTTACAGATGATTAAGGCTAAA 600
DB 181 ValThrSerLeuGlyArgArgIlePheSerLysAsnPheTyrAsnGlyLeuGlnAlaLys 200
QY 601 TCATTAGAGAAATTATTTAGAGAGAAAAAATCTTTCTTAAACCTTTCTTGACACCGCAG 660
DB 201 SerLeuGlnLysAsnLysIleGlnLysLysLeuSerLysProPhePheArgProProGln 220
QY 661 AGATTACCTGATGAGAGAGATGATTTTGGCTGACCAAGACGCGCTTAAATGAGAA 720
DB 221 ArgLeuProAspGlyArgIleGlyTyrLeuAlaGlyProThrGlnLysAlaProLysTyrPag 240
QY 721 GTGAGTTTAAAGAACTTAAATAACAATCTAGAGATGATTTTCTTATATGAGAGG 780
DB 241 ValSerPheLysGlnLysLysAsnLysLysSerArgAsnGlyPheSerAsnMetGlnGly 260
QY 781 GCTGCAAAACAAAGATATGTCATTTATTAAGAGCTACAAAGGTAACGCTCCACAG 840
DB 261 AlaAlaLysGlnLysTyrSerSerPheIleLysGlnValGlnLysGlnLysAlaProGln 280
QY 841 ACAGAGAGGAAAGATTTGCTACAGCAGTGCAGAGTAACCTGAGAAATTTGCGCAATAT 900
DB 281 ThrAlaAlaLysSerIleGlyThrAlaSerGlySerAsnLeuGlnLysLeuProAsn 300
```

```
QY 901 TTATATAGTGTGAGGCTTAAGCCAAAAAGACAGGGTAACCTTACTCAAAATGATATCTGAC 960
DB 301 LeuTyrSerValArgLeuSerGlnLysAspArgValThrPheThrGlnAsnAspThrAsp 320
QY 961 AATACAAATGACGCTTCATAGTGTGGAACCTCAATTTAAATATA 1005
DB 321 AsnThrMetThrValHisSerValGlyThrHisTyrLysAsnIle 335

RESULT 2
US-10-617-962-3
; Sequence 3, Application US/10617962
; Publication No. US20040055036A1
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; APPLICANT: EAST, Peter David
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Phocor
; TITLE OF INVENTION: Luminescens
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/10/617,962
; PRIOR FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: US/09/463,048A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-617-962-3

Alignment Scores:
  Pred. No.:      2,99-116      Length:      368
  Score:          1276.00        Matches:     368
  Percent Similarity: 80.17%    Conservative: 36
  Best Local Similarity: 70.11%  Mismatches:  47
  Query Match:     73.54%       Indels:      24
  DB:              15           Gaps:         3

US-10-617-962-2 (1-1008) x US-10-617-962-3 (1-368)

QY 4 GTTATACATTAACAACCTGATGATGA-----AGTGATATCCACC 45
DB 11 ValIleGlnLeuThrProaspAspArgValThrProaspLysGlyGlnTyrGlnPro 30
QY 46 GTTGAAGGCAAAATGACAGAGATATGATGATGATTAACCTTAAAGCAAAACAGATGAG 105
DB 31 ValGlnLysGlnIleAlaGlyAspIleLeuArgValLeuGlnLysGlnThrAsnGln 50
QY 106 GGTCAATACGATCATATGAGAAATTAATGAGCTTAAGAAATAATATTAGCTTACGCT 165
DB 51 SerHisThrGlyLeuTyrGlyIleAlaTyrArgAlaLysLysValIleIleAlaTyrAla 70
QY 166 TTGGCTGTAAGTGAATTCATATGATGATCTTAAACCTTCGATGACTTATTAAGATATA 225
DB 71 LeuAlaValSerGlyIleHisAsnValSerGlnLysProGlnLysPheTyrLysAsnLys 90
QY 226 GAGACTGCTGAGAGAAATTTATCAAGAAATAATGTCATACTTTTCATCTGCACTATTAGGT 285
DB 91 AspAsnThrGlnArgIleTyrGlnLysIleTyrMetSerAsnLeuLeuSerAlaLeuLysGly 110
QY 286 GAAATGCTGATCAAAATTTCTTAAGATATGCAAAATGCTTTTATTAAGAAATGATCGAT 345
DB 111 GlnAsnGlyAspGlnIleSerLysAspMetAlaAsnAspPheThrGlnAsnGlnLysGln 130
QY 346 TTTGAAGGCAATATCTCAAAACATTTGGAATGATGCTTCGAGCTTGAATAATTAACCATTTG 405
DB 131 PheGlyGlyGlnArgLysLeuLysAsnThrTyrAspIleProaspLeuGlnLysLeuLeu 150
```



```

QY 406 AGTCTTATTCAGATGACGATTAATTTATTAAGCACTAATTTTCTCTGACAGAAATT 465
    |||
    |||
    |||
Db 151 GUAAPTyrSerAspGluAspLysLeuLeuValLeuTyrPhePheAlaSerGlnIleu 170
    |||
    |||
    |||
QY 466 CCAGTGGAGAAATCAACATCAATGCCGACAGATTTTAAATTAATGATTTCTTA 525
    |||
    |||
    |||
Db 171 PrometGluAlaAsnGlnGlnSerAsnAlaAlaAsnPhenylValIleAspPheLeu 190
    |||
    |||
    |||
QY 526 TTTACCTTAATCGCTGCTAATCTTCACTGGAGAGAGATTTTTCGAAATTAATTAAGAT 585
    |||
    |||
    |||
Db 191 LeuIleuSerAlaValThrSerLeuGlySerAlaGlyIlePheSerLysAsnPhenylVal 210
    |||
    |||
    |||
QY 586 GGATTAGAGCTTAATCATTAAGATTAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 645
    |||
    |||
    |||
Db 211 GlyLeuGlnThrLysSerLeuGlnLysThrIleGlnAlaGlyLysLeuSerLysProPhe 230
    |||
    |||
    |||
QY 646 TTTGACCAACCGCAGAGATTAATCTGATGACAGAGATTAATGATTTGCTGACCAAGAGAA 705
    |||
    |||
    |||
Db 221 PheArgProGlnLysLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrLys 250
    |||
    |||
    |||
QY 706 GCGCTAA----- 714
    |||
    |||
    |||
Db 251 AlaProLysLeuProThrThrSerSerThrAlaThrThrSerThrAlaAlaSerSerAsn 270
    |||
    |||
    |||
QY 715 TGAGAGAGAGATTTTAAGAACTTAATAATTAATAATCTAGAGATTAATTAATTAATTAATG 774
    |||
    |||
    |||
Db 271 ThrArgValSerLeuGlnLysLeuArgAsnProSerArgAsnThrPheMetLysMet 290
    |||
    |||
    |||
QY 775 GAAGGGCTGCAGAAACAAAGTAATGTTCAATTAATAAGAGTACAAAGGTAACGCT 834
    |||
    |||
    |||
Db 291 AspSerAlaAlaLysArgLysThrSerSerPheIleLysGlnValGlnLysGlnLysAsp 310
    |||
    |||
    |||
QY 835 CCAGCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 894
    |||
    |||
    |||
Db 311 ProAlaGlnAlaAlaAlaSerIleGlyThrLysSerGlySerAsnPhenylValLysLeuGln 330
    |||
    |||
    |||
QY 895 ---AATAATTAATTAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951
    |||
    |||
    |||
Db 331 GlyArgAspLeuTyrSerIleThrLeuSerGlnGlnLysArgValThrPheSerLysLeu 350
    |||
    |||
    |||
QY 952 GATATTCGACAAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1005
    |||
    |||
    |||
Db 351 AsnThrArgGlnIleMetGlnIleGlnSerValGlyThrHisLysGlnAlaSer 368
    |||
    |||
    |||

```

```

RESULT 3
US-10-437-963-149602
; Sequence 149602, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 149602
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_49919C.1.pcp
US-10-437-963-149602
Alignment Scores: 0.0968 Length: 652
Pred. No.:

```

```

Score: 110.00 Matches: 81
Percent Similarity: 33.25% Conservative: 56
Best Local Similarity: 19.66% Mismatches: 104
Query Match: 6.34% Indels: 171
DB: 16 Gaps: 23
US-10-617-962-2 (1-1008) x US-10-437-963-149602 (1-652)
QY 127 ATGGAATATCGAGCTTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 174
    |||
    |||
    |||
Db 1 MetAspLeuLysThrIleLysIleLeuLeuArgTyrProArgGlySerAspGlnTyrLeu 20
    |||
    |||
    |||
QY 175 AGTGATTCATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 234
    |||
    |||
    |||
Db 21 AlaGlyValHisGlyPheLeuGlnPheAla-----TyrLysAspLysSerGlnAsp 37
    |||
    |||
    |||
QY 235 GAGAGATTTATCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 294
    |||
    |||
    |||
Db 38 ThrLysIleTyrCysProCysGlnThrCysValHisThrMetLeuLeuSerLysAsn--- 56
    |||
    |||
    |||
QY 295 GATCAAAATTTCTAAAGATGAGCAAAATGATTTTATTAATG-----AATGACTGATTTT 348
    |||
    |||
    |||
Db 57 AspValTyrAspArgLeuValCysAsnGlyMetLeuGlnSerTyrAspGlnThrAspPhe 76
    |||
    |||
    |||
QY 349 GAAGGT-----CAATATCTCAAAACATTTGGAATGTT 381
    |||
    |||
    |||
Db 77 HisGlyLysSerSerGlnGlnGlnThrArgAsnGlnGlnProGln----- 91
    |||
    |||
    |||
QY 382 CTTGAGCTTGAATAATTAACATTTGAGTCTTAATCA---GATGACGATTAATTAATTAATGA 438
    |||
    |||
    |||
Db 92 ProHisAsnGlnAsnMetArgAlaAsnMetHisGlnLeuIleAsnAspAlaPheArgThr 411
    |||
    |||
    |||
QY 439 CTATATTTTCTCTGACAGAAATTCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
    |||
    |||
    |||
Db 112 ValTyr-----AspAspValProMetSerAspHisAlaAspSerProCysThr 127
    |||
    |||
    |||
QY 490 -----AATGCCGAGAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 519
    |||
    |||
    |||
Db 128 HisThrAspGlyProAsnLeuGlnAlaGlnAlaPheTyrLysLeuValLysAspSerGln 147
    |||
    |||
    |||
QY 520 -----TTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 531
    |||
    |||
    |||
Db 148 LysProLeuTyrAspGlyCysGlnLeuSerHisLeuSerLeuLeuValLeuLeuPheAsn 167
    |||
    |||
    |||
QY 532 TTATCTGCTGAATCTTCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
    |||
    |||
    |||
Db 168 MetLysSerMetAsnLysTyrPheSerAspLysSerPheGlyAspLeuLeuAspIleLeuHis 187
    |||
    |||
    |||
QY 571 -----AAAACTTTTAACAT----- 585
    |||
    |||
    |||
Db 188 MetAlaIleProAsnGlnLysGlnLeuProLysAsnPhenylValAlaLysLysValVal 207
    |||
    |||
    |||
QY 586 -----GGATTAGAG----- 594
    |||
    |||
    |||
Db 208 SerLysPheGlyLeuAspTyrGlnGlnIleHisAlaCysProAsnAsnCysGlnLeuPhe 227
    |||
    |||
    |||
QY 595 -----GCT 597
    |||
    |||
    |||
Db 228 TrpLysAspLysValAsnAspAspPheCysSerIleCysLysAlaSerArgTyrLysGly 247
    |||
    |||
    |||
QY 598 AAATCATTAAGAGAAATTAATTT-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
    |||
    |||
    |||
Db 248 LysLysProGlnThrLysLeuThrLysSerLysLysValAlaThrProSerLysValLeu 267
    |||
    |||
    |||
QY 631 -----CTTTCT 636
    |||
    |||
    |||
Db 268 ArgTyrPheProIleLysAspArgLeuLysArgLeuPheMetCysArgGlnThrAlaSer 287
    |||
    |||
    |||
QY 637 AAACCTTTCTTTCAGACACCGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 696
    |||
    |||
    |||
Db 288 LeuProAlaGlyThrPheAspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnHis 304
    |||
    |||
    |||
QY 697 CCACAGAGAGCGCTTAATTAATGAGAG-----GTAGGTTTAAAGAACTTAATAATAATAC 747
    |||
    |||
    |||

```

Db	305	ProThrAspSerProAlaTrpLysAsnPheAspGluArgPheIysGlnPheValSerAsp	324
Qy	748	AAATGTAGAAAT-----GGAATTTCTTAATATGAAAGGGGCT	763
Db	325	---SerArgAsnIleArgPheGluLeuAlaThrAspGlyPheAsnProTyrGluGlyLys	343
Qy	784	GCAAAACAAAGCATATAGTTCA-----TTATATAAAGAGGTACAAA---AGGGTAACGCT	834
Db	344	SerTyrAspAsnIleuGlnIleArgLeuAspLeuIleYsGluMet-AsnIleArgGluGluLe	363
Qy	835	CCACAGACAGCAGCGCA-----AAAGATATGTATACAGCCAGTGCACATAACCTTGAA	885
Db	363	uHisProGlnGlnArgThrSerSerIysPheTyrLeuProAlaSerPheThrMetSe	383
Qy	886	AAAT-----TGCCGATTAATT	901
Db	383	rlsSerGluYsGlnLeuPheIysValLeu	394

```

RESULT 4
US-10-425-115-293701
; Sequence 293701, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 293701
; LENGTH: 1581
; TYPE: prt
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_30939C.1.pep
US-10-425-115-293701

```

Alignment Scores:	
Pred. No.:	0.657
Score:	102.50
Percent Similarity:	37.15%
Best Local Similarity:	22.60%
Query Match:	5.91%
DB:	17
Gaps:	17
Length:	1581
Matches:	73
Conservative:	47
Mismatches:	112
Indels:	91
	17

US-10-617-962-2 (1-1008) X US-10-425-115-293701 (1-1581)

CY		103	GAGGGTCATPACGAMCATATGAGAAATGAATA-----	125
Db		869	GtnglytyrThrglnIlectugllyLysaspheGlyglutrhreAlapovalAlaarg	888
OY		136	---CGAGCTAAGAAAAATAATTATTAGCTTACGCCTTGCTGTAAAGTGTATTCAATANGTA	192
Db		889	LengluAlalleargIleuleuAlatryAlacAlala-----HisAsnile	904
CY		193	TCTAACTCCCGATGACTAT-----TATTAAGAATTAAGAGACTGCTGAGAGAAAT	243
Db		905	LysLeuTyrglmMechaspValylsSerAlaphelLeuasnmglylysIleSerGluLeuval	924
OY		244	TATCAAGAA-----TATATGTCTAATCTT	267
Db		925	TyrValglungInproProglypheGuaapproLyasgrProasnHisValPhelyLeu	944
CY		268	TCATCTGCACATATTAGGTGAATAATGSGTAGATCCAATTCTTAAGATATNGCAAAATGCTTT	327
Db		945	SerlysalaleuTyrltyl-----LeuylsglnAlaproArgAlatrp	958
OY		328	TATAAGATGAACGTGATTTTGAAGGTCAATATCCTCAAAAACAATTTGGAAATGTT---CCT	384

[illegible]

RESULT 5  
US-10-087-464-10  
; Sequence 10, Application US/10087464  
; Publication No. US20030059436A1

```

/ APPLICANT: Chishti, Aghar
/ APPLICANT: Oh, Steven
/ APPLICANT: Liu, David
/ APPLICANT: Goel, Vikas
/ APPLICANT: Li, Xuetong
/ TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
/ FILE REFERENCE: S137/7019
/ CURRENT APPLICATION NUMBER: US/10/087,464
/ CURRENT FILING DATE: 2002-03-01
/ PRIOR APPLICATION NUMBER: US 06/272,930
/ PRIOR FILING DATE: 2001-03-02
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 10
/ LENGTH: 1639
/ TYPE: PRT
/ ORGANISM: Plasmodium falciparum
US-10-087-464-10

Alignment Scores:          0.832          Length:          1639
Pred. No.:

```



```
Db      1263 ThrLeuPheThrLysArgIleGlyLysAspLeuPheValCysGlnIleTyrValAspAsp 1302
QY      424 GATTAATATTATGACATATATTTCTCTGTACAGAAATTCACGTGAGGAAATCA 483
Db      1303 -----IleIlePheGlySerThrAsnGlnLeuPheCysGlnGlu----- 1315
QY      484 CAATCAATGCCGACAGATTTTAAATTAATTGATTTCTTATTTACCTTATCTGCTGA 543
Db      1316 -----PheGlyLysMetMetSerLysGlnPheGlnIleuSerMetIle 1329
QY      544 ACTTCACGTGGAGAGATTTTTCAAAACCTTTACATGAGATTAGAGCTTAATCA 603
Db      1330 GlyGluLeu-----SerPhePheLeuGlnIleuGlnIleuSerGln 1342
QY      604 TTAGAGAAAT-----TATATTGAGAGAAAAAACTTTCTAACCTTTCTTGA----- 651
Db      1343 LeuLysAspGlyIlePheIleSerGlnSerLysTyrLeuLysAspMetLeuLysPhe 1362
QY      652 -----CCACCGCAGAGATTACCTGATGCGAAGATAGCTTATTTG----- 690
Db      1363 GlyLeuGluAsnAlaLysProIleLysThrProMetAlaThrAsnGlyHisLeuAspLeu 1382
QY      691 -----GCTGACCAACAGAAAGCGCCTAATGAGAGTGAAGTTTAAAGA 735
Db      1383 AspGluGlyValAlaCysGlnLysSerAspSerLeuGln-----GlnHis 1397
QY      736 CTTAAATATACAAATCTAGGAATGGATTTTCTAATATGAGAGGCGTGCACAAACAAAG 795
Db      1398 SerLeuAsnValSerSerSerSerArgLysPheGly-----GlyLeuArgLysLysLeu 1414
QY      796 TATAGTCTATTATTAAGAGGTACAAAAGGTAAC-----GCTCCACAGACA 843
Db      1415 AlaGlyArgPheLysSerLysArgProArgLysAspAspAspLysPyrThrProThrThr 1434
QY      844 GCAGCGAAAGATATTGTGTACAGCAGCAGTGGCAGT 876
Db      1435 AspSerGluAlaGlnSerSerAlaGlyGlyThr 1445

RESULT 7
US-10-425-115-344697
; Sequence 344697, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 344697
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_77529C.1.pcp
US-10-425-115-344697

Alignment Scores:
Pred. No.: 0.763 Length: 297
Score: 100.00 Matches: 58
Percent Similarity: 34.16% Conservative: 25
Best Local Similarity: 23.87% Mismatches: 76
Query Match: 5.76% Indels: 84
DB: 17 Gaps: 12

US-10-617-962-2 (1-1008) x US-10-425-115-344697 (1-297)
QY      145 AAAATAATATTAGCTTACGCT-----TTGGCTGTAGAGTGATTCATAATGATATCT 195
```

```
Db      95 LysValThrLeuAlaIleAlaIleAspAspLeuCysLeuIleGlyPheGlnAsnLaser 114
QY      196 AAATCTCTGATGACATATATTAAGAAATTAAGAGACTGCTGAGAGATTTATCAAGAAATAT 255
Db      115 -----AspHisTrpTyrLys----- 119
QY      256 ATGTCTAATCTTTCACTGTCACATTAATAGGTAAATAGT-----GAT 297
Db      120 PheAsnGlyGlnSerSerPheLysGlyLeuProGlyAlaThrValLeuProIleArg 139
QY      298 CAATTTCTAAAGATATGCGCAAGTGTTTTATAAGATGAACGTGATTTTGAAGGTCA 357
Db      140 GlnAsnTyrGlnAspLeuIleLysGlyHis----- 149
QY      358 TATCTCAAAACATTTGGAATGTTCTGAGCTGGAATAATTAACATTAATGATGCTTATTC 417
Db      150 -----AlaAsnLeuTrpLysValProLeuGlyLysSerAlaIleHisAla----- 165
QY      418 GATGACGATTAATTTATGACACTATATTTTCTCTGTA---CAGGAATTCACGTGAG 474
Db      166 -----ThrLysGlnLeuAlaThrTyrAspArgAlaValThrProAspSerGluLeuLys 183
QY      475 GAAAT-----CAACATCAATGCCGAGATTTTAAATTAAT 516
Db      184 AspGlyLeuAlaArgPheValAlaMetCysGlnGlyMetArgPheArgSerIleArg 203
QY      517 GATTTCTTATTTACCTTATCTGCTGTAACTTCACTGGAAGAGATTTTTCAAAAC 576
Db      204 AspMetPheSerSerLeuSer-----GlyAsnAsnTrpGluGlnGluThr 218
QY      577 TTTTACAATGATTTAGAGCTTAATCATTAAGAAATTAATGACAGAAAAAACTTTCT 636
Db      219 PheIleThrGluLeuGlnAlaLysSerValAlaTyrTrpSerGln----- 233
QY      637 AAACCTTTCTTTCACACGCGACAGAGATTACTGATGCGAATTAAGTATTATGGCTGA 696
Db      234 -----LeuProProArgArgLeuArgSerAlaArgAlaGlyArgLeuProGly 249
QY      697 CCACAGAGAGCGCTTAATGAGAGTGAAGTTTTAAAGAACTTAATTAACAAATCTAGG 756
Db      250 -----GlyProLysTrp----- 253
QY      757 AATGATTTTCTAATATGAAAGGCGCTGCACAAAAGTAAGTCTTATTAATAAGAG 816
Db      254 -----GlyAlaValAspGlyArgTyrAsnSerMetAlaLysHis 266
QY      817 GTACAAAG 825
Db      267 ValGlnGlu 269

RESULT 8
US-10-369-493-19786
; Sequence 19786, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19786
; LENGTH: 367
; TYPE: PRT
; ORGANISM: No. US20030233675A1loc punctiforme
```





QY 682 GGTATTGCTGCAACAGAGCGCTAAATGAGATGATTAAAGACTTAA 741  
 Db 967 yllegrtprtlegrlnasprthleutrpProsmgly--ValpPheThArgrleuasp 986  
 QY 742 AAT-----AACAAATCTAGAAATGATTTTCTAAT 771  
 Db 987 GlyTYrGlnGlyAsnAlaGlyProserGlnPheapLySHAsProserGlySerAlaSn 1006  
 QY 772 ATG--GAAGGCGCTGCAACAAAGTATAGTTCTATT----- 807  
 Db 1007 GlnAlaSerGlyAsnArgLyAspSerAlaSerSerPheGluGlnGlnLeuGlnAlaSer 1026  
 QY 808 -----ATTAAGAGGTACAAAAGGTAAAGCTCCACAGCAGCAGAAA 852  
 Db 1027 ArgAsnAlaSerGluValValLeuLeuLeuGlyGlyThrProThrLeuValSer 1046  
 QY 853 AGTATTGGT 861  
 Db 1047 llellegly 1049

RESULT 11  
 US-10-282-122A-54429  
 ; Sequence 54429, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zykend, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Foreyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282.122A  
 ; PRIOR FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See file wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: Patent version 3.1  
 ; SEQ ID NO 54429  
 ; LENGTH: 580  
 ; TYPE: PRF  
 ; ORGANISM: Campylobacter jejuni  
 ; US-10-282-122A-54429  
 Alignment Scores: 1.42 Length: 580

Score: 98.00 Matches: 74  
 Percent Similarity: 35.99% Conservative: 48  
 Best Local Similarity: 21.83% Mismatches: 135  
 Query Match: 5.65% Indels: 82  
 DB: 15 Gaps: 14

US-10-617-962-2 (1-1008) x US-10-282-122A-54429 (1-580)

QY 181 ATTCAATATGATATCAAACTTCGATGACATATATTAAGATTAAGAGACTGAGAGA 240  
 Db 103 LeuysAsnValLeuArgLeuAspMetCAspPheLySArgTYrArgSerGlyGlnLeu 122  
 QY 241 ATTATCAAGATATATATGCTTAATCTTTCATCTGCACATATTAGGAAATGAGATCA 300  
 Db 123 ValSerArg-----CysThrAsnAspIleAsnAlaLeu-----GlnSer 135  
 QY 301 ATTTCTAAAGATATGCGCAATATGTTTATTAAGATCAAGATTTTGAAGTCATAT 360  
 Db 136 IleValSerAsnIleIleProAspPhePheArgGlnLeuThrAlaIleGlyLeuLeu 155  
 QY 361 CCTCAAAACATTTGGAAATGTTCTGAGCTTGA----- 393  
 Db 156 AlaValValLeuTYrGlnSerProThrLeuAlaPhePheAlaLeuValIleLeuProLeu 175  
 QY 394 -----AATAACCATTTGAGTCTTAT----- 414  
 Db 176 AlaIlePheProLeuValTrpPheAlaIleValLeuValLeuValArgAsnIleGln 195  
 QY 415 -----TCAGATGACATTAATATTTATGACATATATTTTCTCTGACAGAAATT 465  
 Db 196 GluThrAsnSerAspLeuLeuSerTYrLeuGlyGlnIlePheSerAsnIleGlyLeuIle 215  
 QY 466 CCACTGAGAAATCAACATCAAAATGCCCAAGATTTTAAATTAATTAATGATTTCTTA 525  
 Db 216 LysAlaAsnAspAsnGluValLysGluSerAspIleAlaValHisAsnAspThrLeu 235  
 QY 526 TTACCTTATCT-----GCTGAACCTTCACTG----- 552  
 Db 236 CysIleValAsnLeuLeuValSerAlaArgIleAspAlaLeuThrSerProLeuMetCAspMet 255  
 QY 553 -----GAGAGGAGATTTTTCAAAA 573  
 Db 256 MetGlySerValGlyValAlaValAlaValIleIleValGlyGlyArgGlnValIle----- 273  
 QY 574 AACTTTTCAATGATGATTAAGAGCTTAATCACTTAAGATTTATATGAGAGAAAAACTT 633  
 Db 274 -----AsnGly-----SerMetSerValGlySerPheIleSerPheValSerAla 288  
 QY 634 TCTAAACCTTCTTTCGACACCGCAGATTAATTAATGAGATTAATGATTTATTTGGCT 693  
 Db 289 LeuPheAlaIleTYrThrProLeuValArgLeu--SerSerLeuTYrGlyLysLeuGln 307  
 QY 694 GAGACCAAGAGAGCG-----CTTAATGAGAGA 720  
 Db 308 GlyAlaValAlaIleAlaSerGluArgThrPheTYrLeuLeuAspLeuGlnProGlnIleLys 327  
 QY 721 GTGAGTTTAAAGAACTTAAATAATCAAACTAGATGATTTTCTTAATATGAGAGG 780  
 Db 328 GlyGlySerLysGluLeuLysAsnIleGlu--LysIleSerPheGluAsnValGlnPhe 346  
 QY 781 GCTGCAAAACAAAGTATGATGTTCACTTTTAAAGAGGTACAAAGGTTAGCGTCCACAG 840  
 Db 347 AlaTYrGluAsnProHisLysSerValLeuLysGlyValAsnPheAspPheValLysGly 366  
 QY 841 ACAGCAGGAAAGTATGTTGTTACGCGCAGTGGCAGTTAACTTGAGAAAATTCGCCAATAT 900  
 Db 367 GlnMetLeuAlaLeuValGlyThr--SerGlyGlyGlyLysSerSerIleIleAsnLeu 385  
 QY 901 TTATATAGTGTAGAGCTTAAGCAAAAGACAGGGTAACCTTATCAAAAATGATCTGAC 960  
 Db 386 LeuMetTYrPheTYrGluLysGlnLysIleLysIleLeuLeuAsnGlnLysAspIle--- 404  
 QY 961 AATACAAAGACGGTTTCATATG-----GTTGGAACCTCATTTATTAATAATATA 1005





```

FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_17949C.1.pep
US-10-425-115-271828

Alignment Scores:
Pred. No.: 2.09 Length: 1770
Score: 97.50 Matches: 71
Percent Similarity: 36.88% Conservative: 47
Best Local Similarity: 22.19% Mismatches: 117
Query Match: 5.62% Indels: 85
DB: 17 Gaps: 16

US-10-617-962-2 (1-1008) x US-10-425-115-271828 (1-1770)

QY 103 GAGGCTCATACGATCATATGGAATGTAAT----- 135
DB 1208 GlnGlyTyrThrArgTlGlnGlyLeuAspPheGlyGluArgPheAlaProValAlaArg 1227
QY 136 ---CGAGCTAAGAAATATATATACCTTACGCTTGGCTGTAAGTGTATTCATATGTA 192
DB 1228 LeuGluAlaIleArgTlLeuLeuAlaTyrAlaCysAla-----HisenIle 1243
QY 193 TCTAACTCTCGATGACTAT-----TATAAGAAATTAAGAGACTGCTGAGAAAT 243
DB 1244 LysLeuTyrGlnMetAspValIleSerAlaPheLeuAlaGlyLysIleSerGluLeuVal 1263
QY 244 TATCAAGA-----TATATGCTTAATCTT 267
DB 1264 TyrValGlnGlnProProAspPheGluAspProLysArgProIleHisValPheLysLeu 1283
QY 268 TCATCTGCATATGATGTAAGAAATGCTGATCAATTTCTTAAGATATGCAAAATGCTTT 327
DB 1284 SerLysAlaLeuTyrGly-----LeuLysGlnAlaProArgAlaTrp 1297
QY 328 TATAAGAAATGACTGATTTTGAAGCTCATATCTCAAAACATTTGGAATGT----- 381
DB 1298 TyrGluArgLeuArgAspPheLeuLeuSerLysAspPheArgTlGlyLysValAspThr 1317
QY 382 ---CCTGAGCTTGAATAATTAACCATTTG-----AGTGTATTATTCAGATGAC 423
DB 1318 ThrLeuPheThrLysIleIleGlyLysAspLeuPheValCysGlnIleTyrValAspAsp 1337
QY 424 GATAAATATATGACTATATTTTCTCTGTACAGAAATCCATGAGGAGAAATCA 483
DB 1338 -----IleIlePheGlySerThrAsnGlnLeuPheCysGlu----- 1350
QY 484 CAATCAAAATGCCGAAGATTTTAAATTAATTAATGATTTTAACTTATCTGCTGA 543
DB 1351 -----PheGlyLysMetMetSerLysGluPheGluMetSerMetIle 1364
QY 544 ACTTACTGGGAGAGAGATTTTCAAAAACTTTTACATGATGATTAGAGCTTAATCA 603
DB 1365 GlyIleLeu-----SerPhePheLeuGlnIleLeuGlnIleLysGln 1377
QY 604 TTAGAGAAAT-----TATATTGAGAGAAATAAATCTTCAACCTTTCTTGA----- 651
DB 1378 LeuLysAspGlyTlPheIleSerGlnSerLysTyrLeuLysAspMetLeuLysPhe 1397
QY 652 ---CCACCGCAGAGATTACTGATGAGGAGAAATGAGTATTTG----- 690
DB 1398 GlyLeuGluAsnAlaLysProIleLysThrProMetAlaThrAsnGlyHisLeuAspLeu 1417
QY 691 ---GCTGACCAACAGAA-----GGCCTTAATGAGAGATGAGTATTT 729
DB 1418 AspGlnGlyAlaSerProArgGlnIleHisLeuLysAlaAlaLysArgTlLeuArgTyr 1437
QY 730 AAAAGAACTTAATAATCAAAATCTAGAAATGATTTTCAATATGAGAGGCTGCAAAA 789
DB 1438 LeuLysTyrThrProAsnIleGlyLeuLysGlyLeuGlyLeuArgLysLysLeuAla 1457
QY 790 CAAAAGTATAGTTCAATTATTAAGAGGTACAAAAGGTAAC---GCTTCCACAGACAGCA 846
DB 1458 GlyArgPheLysSerLysArgProArgValAspAspAlaAspTyrAsnProThrThrAsp 1477

```

```

QY 847 GCGAAAGTATTTGGTACAGCCAGTGGCAGT-----AACCTGAAAAATTCGCGAAT 897
DB 1478 SerGlnAlaGlnSerLeuAlaGlyGlySerValSerMetAspThrGlnAspValProHis 1497

RESULT 14
US-10-425-115-249308
; Sequence 249308, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: US/10/425, 115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 249308
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_158956C.1.pep
US-10-425-115-249308

Alignment Scores:
Pred. No.: 1.97 Length: 878
Score: 97.00 Matches: 74
Percent Similarity: 34.75% Conservative: 49
Best Local Similarity: 20.90% Mismatches: 137
Query Match: 5.59% Indels: 94
DB: 17 Gaps: 14

US-10-617-962-2 (1-1008) x US-10-425-115-249308 (1-878)

QY 19 CCTGATGATGAAGTATATCCACCGCTTGAAGAAATATGACAGAGATATAGTACT 78
DB 523 ProAspIleLysThrSer---SerProIleGlnLysGlnGlyAlaGlnLeuTyrThrArg 541
QY 79 ---ATACATAACTTTTAAACAA----- 96
DB 542 SerMetPheLeuLysPheGlnGlnLeuLeuIleCysAlaThrAlaTyrThrAlaGluMet 561
QY 97 ---ACAGATGAGGCTATACAGCATATATGAAAT----- 129
DB 562 ValLysGlnGlnGlyAsnAlaSerIleTyrThrValAlaIleSerGlnGlySerGluLys 581
QY 129 ----- 129
DB 582 SerValThrValGlnPheCysSerSerGlySerSerAlaThrCysSerCysArgMetPhe 601
QY 130 GAATATGCAAGCTTAAGAAATATATATAGCTTACGCTTGGCTGTAAGTGTATTCATAAT 189
DB 602 GlnTyrPheGly-----IleValCysArgHisIleLeuThrValPheGlyValArgGly 619
QY 190 GTATCTAAACTCTCGATGACTATATTAAGATTAAGATAAGAGACTGCTGAGAGAAATTTATCA 249
DB 620 ValSerAlaLeuProSerHisTyrPheValLysArgTlPhe----- 633
QY 250 GAATATATGCTTAATCTTTCATCTGCACTATTTAGGTGAATAATGGTATCAATTTCTAAA 309
DB 634 ---LysAsnAlaLeuAspArgSerLeuGlnLysAsnValAspGlnValSerArg 650
QY 310 GATATGCAAAATGCTTTTATTAAGATGAACCTGATTTTGAAGGCTCAATATCTCAAAAC 369
DB 651 ValGlnProLysGlnGlnGlnArgSerSerAlaGluAspGlyGlnGlnSerPro----- 668
QY 370 ATTTGAATGTTCTGACCTTGAATAATCAACATTTAGTGTATTTCAATGACAGATAAA 429
DB 669 ThrTrpArgTyrAsnSerLeuCysHisGlnAlaLeuArgTyrValAlaGlnGlnGlyAlaSer 688

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_nzp model

Run on: November 21, 2004, 07:40:10 ; Search time 92 Seconds

(without alignments)  
7860.851 Million cell updates/sec

Title: US-10-617-962-2  
Perfect score: 1735  
Sequence: 1 atgcttatacattacacc.....ctcattataaataatgta 1008

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2002273 seqs, 35872929 residues

Total number of hits satisfying chosen parameters: 4004546

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+npz\_model -DEV=xlp  
-O=/cgn2.1/USPTO\_spool\_p/US10617962/runat\_17112004\_163921\_9743/app\_query.fasta\_1.1159  
-DB=A\_Geneseq\_23Sep04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blowum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10617962 @CGN 1.1.154 @runat\_17112004\_163921\_9743 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: A\_Geneseq\_23Sep04:\*  
2: geneeqp1980s:\*  
3: geneeqp1990s:\*  
4: geneeqp2000s:\*  
5: geneeqp2001s:\*  
6: geneeqp2002s:\*  
7: geneeqp2003as:\*  
8: geneeqp2003bs:\*  
9: geneeqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1734	99.9	335	2	AAM97811 Photorhab
2	1716	98.9	340	6	ABM67356 Photorhab
3	1276	73.5	368	2	AAM97810 Xenorhabd
4	940.5	54.2	278	2	AAR68961 Xenorhabd
5	108.5	6.3	1654	1	AAP50777 Sequence
6	108	6.2	940	5	ABR47334 Listeria
7	105.5	6.1	816	6	ADA35390 Acinetoba
8	105.5	6.1	1270	7	ADK83729 Bacillus
9	104.5	6.0	2197	8	ADK16573 Nanorarcha
10	104	6.0	583	6	ABR63439 S suis ar

11	101.5	5.9	1639	2	AAM54145
12	101.5	5.9	1639	5	AAE29345
13	101	5.8	566	6	ABR53129
14	101	5.8	566	7	ADK62560
15	100	5.8	2404	8	ADP25426
16	98.5	5.7	616	6	ABU48940
17	98	5.6	580	6	ABU26505
18	98	5.6	954	6	ABU35684
19	98	5.6	1005	5	ABU35998
20	98	5.6	1413	4	ABR62045
21	97.5	5.6	511	2	AAV37593
22	97	5.6	1714	3	AAU18275
23	96.5	5.6	685	4	AAU37884
24	96.5	5.6	685	6	ABU02144
25	96.5	5.6	685	6	ABP81632
26	96.5	5.6	685	6	ABU46194
27	96	5.5	384	7	ADU19754
28	96	5.5	512	4	ABR71449
29	96	5.5	623	6	ABR53842
30	96	5.5	623	7	ADK64956
31	96	5.5	645	2	AAH86966
32	96	5.5	879	6	ABU26364
33	96	5.5	912	2	AAU94029
34	96	5.5	3168	7	ABO67550
35	95.5	5.5	303	6	ABU23912
36	95.5	5.5	360	6	ABU26457
37	95.5	5.5	680	8	ADK46676
38	95.5	5.5	915	2	AAH94700
39	95.5	5.5	1791	8	ADP25441
40	95	5.5	500	5	ABP29236
41	95	5.5	577	4	ABG05216
42	95	5.5	577	4	ABG15756
43	95	5.5	577	4	ABG29786
44	95	5.5	732	4	ABG13490
45	95	5.5	1048	4	ABG21867

## ALIGNMENTS

RESULT 1	AAW97811	standard; protein; 335 AA.
AC	XX	AAW97811;
DT	XX	21-MAY-1999 (first entry)
DE	XX	Photorhabdus luminescens insecticidal toxin protein PIV6tox1.
KW	XX	Toxin; PIV6tox1, biological control; insecticide.
OS	XX	Photorhabdus luminescens.
PN	XX	WO9903328-A1.
XX	XX	28-JAN-1999.
XX	XX	17-JUL-1998; 98WO-AU000562.
XX	XX	17-JUL-1997; 97AU-00008088.
XX	XX	PR
XX	XX	PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX	XX	FI
XX	XX	FI
XX	XX	WIPI; 1999-131733/11.
XX	XX	N-PESDB; AAX07350; AAX07352.
XX	XX	PT New insecticidal toxin genes - extracted from Xenorhabdus nematophilus
XX	XX	PT A24 and Photorhabdus luminescens.
XX	XX	PS Claim 6; Page 29-30; 48pp; English.
XX	XX	

CC This is the amino acid sequence of the insecticidally-active toxin  
 CC protein, termed PIV6toxi, of *Photobabidus luminescens* strain V16/1. It  
 CC was deduced from the coding region (see AAX07350) of the PIV6toxi gene,  
 CC which was isolated from a DNA library using the *Xenorhabdus nematophilus*  
 CC toxB4 gene (see AAX07349) as probe. The invention relates to the  
 CC identification and isolation of polynucleotide molecules encoding a new  
 CC class of protein insecticidal toxins which are produced by bacteria of  
 CC the genera *Xenorhabdus* and *Photobabidus*. The polynucleotide molecules may  
 CC be incorporated e.g. into insect-specific viruses (e.g. entomopox and  
 CC nuclear polyhedrosis viruses), bacteria (e.g. *Gracilicutes*, *Firmicutes*,  
 CC *Tenericutes* and *Mendosicutes*), protozoa, yeast and plants (particularly  
 CC cereals such as wheat and barley, vegetables such as tomato and potato,  
 CC fruit trees such as citrus and apple, tobacco and cotton) for control of  
 CC pest insects

CC Sequence 335 AA;

CC Alignment Scores:

Pred. No.:	1,22e-168	Length:	335
Score:	1734.00	Matches:	335
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.94%	Indels:	0
DB:	2	Gaps:	0

US-10-617-962-2 (1-1008) x AAW97811 (1-335)

QY 1 ATGTTATACATTAATTAACCTGATGATAGAGTGTATCCACCCGTTGAAAAAGCAATA 60  
 DB 1 MetValIleGlnLeuThrProAspAspArgSerGlyTyrProProValIleuLysGlnIle 20  
 QY 61 GCAGAGATATATGTAACGTAATTAACCTTAAGCAACAGATGAGGCTCATACAGCATCA 120  
 DB 21 AlaGlyAspIleValAlaGlnLeuAsnPhenylGlnThrAspGlnGlyHisThrAlaSer 40  
 QY 121 TATGGAATTTGATATCGAGCTAAGAAATTAATTTAGCTTGGCTTGAAGTGT 180  
 DB 41 TyrGlyIleGlnTyrArgAlaLysLysIleIleuAlaTyrAlaLeuAlaValSerGly 60  
 QY 181 ATTCAATATGATCTTAACCTTCCTGATGACTATTTATAGAAATAAGAGACCTGAGAGA 240  
 DB 61 IleHisAsnValSerIleuLeuProAspAspTyrTyrLysAsnLysGlnThrAlaGlnArg 80  
 QY 241 ATTATCAAGATATATATGCTTAATCTTTGATCTGCACTATTAAGTGAATAATGGTGATCA 300  
 DB 81 IleTyrGlnGlnLysTyrMetSerAsnLeuSerSerAlaLeuLeuGlyGlnLysGlnArg 100  
 QY 301 ATTCTTAAGATATATGCTTAATCTTTGATCTGCACTATTAAGTGAATAATGGTGATCA 360  
 DB 101 IleSerLysAspMetAlaAsnGlyPheTyrLysAsnGlnLeuAspPheGlnGlnTyr 120  
 QY 361 CCTCAAAACATTTGGAATGCTTCCTGAGCTTGAATAATCAATTGAGTCTTATTCAGAT 420  
 DB 121 ProGlnAsnIleTyrAsnValProGlnLeuGlnAsnLysProLeuSerAlaTyrSerAsp 140  
 QY 421 GAGCAATTAATTAATGACCTATATTTTCTCTGACAGCAATTCACCTGAGAGAAAT 480  
 DB 141 AspaAspLysLeuLeuAlaLeuTyrPhePheSerValGlnGlnIleProLeuGlnGlnAsn 160  
 QY 481 CAACATCAATCAATGCGCAGATTTTAAATTAATTAATGATTTCTTAACTTAACTGAT 540  
 DB 161 GlnGlnSerAsnAlaAlaIleArgPhePheLysLeuIleAspPheLeuSerThrLeuSerAla 180  
 QY 541 GTAATCTTCACTGGAAGAGAGATTTTTCATCAAAAACCTTTACATGATGATGAGCTTAA 600  
 DB 181 ValThrSerLeuGlyAlaArgArgIlePheSerLysAsnPhenylAsnGlyLeuGlnAlaLys 200  
 QY 601 TCATTAAGAGATTAATTAATGAGAAAAAACTTTCTTAACTTTCTTGGACCCGCGAG 660  
 DB 201 SerLeuGlnLysAsnTyrIleGlnArgLysLysLeuSerLysProPhePheArgProProGln 220  
 QY 661 AGATTACCGATGAGAGATAGTTATTTGGCTGACCAACGAGACCGCTTAATGGAATA 720

DB 221 ArgLeuProAspGlyArgIleGlyTyrIleuAlaGlyProThrGlnAlaProLysTyrArg 240  
 QY 721 GTGAGTTTAAAGAACTTAAATAATCAATCTAGAAATGATTTTCTTAATTAATGAGAGCG 780  
 DB 241 ValSerPheLysGlnLeuLysAsnLysSerAlaGlnGlyPheSerAsnMetGlnGly 260  
 QY 781 GGTGCAAAACAAAGATATGTTCAATTTATTAAGAGGTTCACAAAGGTTACGCTCCACAG 840  
 DB 261 AlaAlaLysGlnLysTyrSerSerSerPheIleuGlnValGlnLysGlyAsnAlaProGln 280  
 QY 841 ACAGCAGCAAAAAGATTTGATACAGCCAGTGCAGTAACTCGGAAAAATTTGCCGAAATAT 900  
 DB 281 ThrAlaAlaLysSerIleGlyThrAlaSerGlySerAsnLeuGlnLysLeuProAsnAsn 300  
 QY 901 TTATATAGTGTAGGCTTAAGCCAAAAAGACAGGTTACCTTTACTCAAAATGATATGAC 960  
 DB 301 LeuTyrSerValArgLeuSerSerLysAspArgValThrPheThrGlnAsnAspThrAsp 320  
 QY 961 AATACAAATGACGCTTCATATGTTGGAACCTCATTTTAAATAATA 1005  
 DB 321 AsnThrMetThrValHisSerValGlyThrHisTyrLysAsnIle 335  
 RESULT 2  
 ABM67356  
 ID ABM67356 standard; protein; 340 AA.  
 AC ABM67356;  
 DT 20-NOV-2003 (first entry)  
 DE Photobabidus luminescens protein sequence #453.  
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 OS Photobabidus luminescens.  
 XX  
 XX  
 XX WO200294867-A2.  
 XX  
 XX PD 28-NOV-2002.  
 XX  
 XX PF 07-FEB-2002; 2002WO-IB003040.  
 XX  
 XX PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 XX DR MPI; 2003-148459/14.  
 XX  
 XX PT Genomic sequence of *Photobabidus luminescens* and encoded polypeptides,  
 XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 PS Claim 2; SEQ ID NO 453; 1205bp; French.  
 CC The invention relates to the isolation of genes and their encoded  
 CC proteins from *Photobabidus luminescens*. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of *P. luminescens*  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than *P. luminescens* and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by *P.*  
 CC *luminescens*. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins

XX Sequence 340 AA;

Alignment Scores:

Pred. No.:	8,59e-167	Length:	340
Score:	1716.00	Matches:	331
Percent Similarity:	99.10%	Conservative:	1
Best Local Similarity:	98.81%	Mismatches:	3
Query Match:	98.90%	Indels:	0
DB:	6	Gaps:	0

US-10-617-962-2 (1-1008) x ABM67356 (1-340)

QY 1 ATGGTTATACAAATTAACCTGATGATGAAAGTGATATCCACCCGTTGAAAGCAATA 60  
 DB 5 MetValIleGlnLeuThrProAspAspArgSerGlyTyrProProValGlnLeuGlnIle 24  
 QY 61 GCAGGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
 DB 25 AlaGlyAspIleValAlaArgIleLeuAsnPhenylSerGlnThrAspGlyHisThrAlaSer 44  
 QY 121 TATGGAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
 DB 45 TyrGlyIleGlnLeuTyrArgAlaAlaArgIleLeuAlaTyrAlaLeuAlaValSerGly 64  
 QY 181 ATTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 DB 65 IleHisAsnValSerLeuLeuProAspAspArgTyrTyrAsnLeuGlnThrAlaGlnArg 84  
 QY 241 ATTATTAAGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 DB 85 IleTyrGlnGlnTyrMetSerAsnLeuSerSerAlaLeuLeuGlnGlnAsnGlyAspGln 104  
 QY 301 ATTCTTAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 DB 105 IleSerLeuAspMetAlaAsnGlyPheTyrIleAsnGlnLeuAspPheGlnGlnTyr 124  
 QY 361 CCTCAAAACATTTGGAATGTTCTCGAGCTTGAATAAACCATTGAGTCTTATTCAGAT 420  
 DB 125 ProGlnAsnIleTyrAsnValProGlnLeuGlnAsnLeuProLeuSerAlaTyrSerAsp 144  
 QY 421 GACGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 DB 145 AspAspLeuLeuLeuAlaLeuTyrPhePheSerValGlnGlnIleProLeuGlnLeuAsn 164  
 QY 481 CAACATCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 DB 165 GlnGlnSerAsnAlaAlaArgPhePheLeuLeuIleAspPheLeuLeuThrLeuSerAla 184  
 QY 541 GTAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 DB 185 ValThrSerLeuGlyArgArgIlePheSerLeuAsnPhenylAsnGlyLeuGlnLeuSer 204  
 QY 601 TCATTAGGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 DB 205 SerLeuGlnLeuSerIleGlnTyrGlySerPheProLeuSerPhePhePheProGln 224  
 QY 661 AGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 DB 225 ArgLeuProAspGlyArgIleGlyTyrLeuAlaGlyProThrGlnAlaProLeuStrArg 244  
 QY 721 GTGAGTTTAAAGAACTTAAATTAACAATCTAGGAATGATGATGATGATGATGATGATGAT 780  
 DB 245 ValSerPheLeuGlnLeuLeuAsnAsnSerAsnGlyPheSerAsnMetGlnGly 264

QY 781 GCTGCAGAAACAAAGTATAGTTCATTATTAAGAGGTACAAAAGGTACGCTCCACAG 840  
 DB 265 AlaAlaGlnGlnTyrTyrSerSerPheIleGlnValGlnGlyAsnAlaProGln 284  
 QY 841 ACAGCAGGAAAGATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 DB 285 ThrAlaAlaGlnSerIleGlyThrAlaSerIleSerAsnLeuGlnLeuProAsn 304  
 QY 901 TTATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 DB 305 LeuTyrSerValArgLeuSerGlnLeuAspArgValThrPheThrGlnAsnAspThrAsp 324  
 QY 961 AATACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1005  
 DB 325 AsnThrMetThrValHisSerValGlyThrHisTyrIleAsnIle 339

RESULT 3

AAW97810  
 ID AAW97810 standard; protein, 368 AA.

AC AAW97810;

DT 17-OCT-2003 (revised)  
 DT 21-MAY-1999 (first entry)

DE Xenorhabdus nematophilus insecticidal toxin protein toxh4.

KM Toxin; toxh4; biological control; insecticide.

OS Xenorhabdus nematophila.

PN WO9903328-A1.

PD 28-JAN-1999.

PF 17-JUL-1998; 98WO-AU000562.

PR 17-JUL-1997; 97AU-00008088.

PA (CSTR) COMMONWEALTH SCI & IND RES ORG.

PI East PD;

DR MPI; 1999-131733/11.

DR N-PSDB; AAX07349; AAX07351.

PT New insecticidal toxin genes - extracted from Xenorhabdus nematophilus

PS A24 and Photorhabdus luminescens.

PS Claim 6; Page 26; 48pp; English.

XX This is the amino acid sequence of the insecticidally-active toxin  
 CC protein, termed toxh4, of Xenorhabdus nematophilus strain A24. It was  
 CC deduced from the coding region (see AAX07349) of the toxh4 gene, which  
 CC was isolated from a DNA library by functional screening using Gallieria  
 CC mellonella fourth instar larvae. The invention relates to the  
 CC identification and isolation of polynucleotide molecules encoding a new  
 CC class of protein insecticidal toxins which are produced by bacteria of  
 CC the genera Xenorhabdus and Photorhabdus (see also AAX07350). The  
 CC polynucleotide molecules may be incorporated e.g. into insect-specific  
 CC viruses (e.g. entomopox and nuclear polyhedrosis viruses), bacteria (e.g.  
 CC Gracilicutes, Firmicutes, Tenacitutes and Mendosicutes), protozoa, yeast  
 CC and plants (particularly cereals such as wheat and barley, vegetables  
 CC such as tomato and potato, fruit trees such as citrus and apple, tobacco  
 CC and cotton) for control of pest insects. (Updated on 17-OCT-2003 to  
 CC standardise OS field)

XX Sequence 368 AA;

Alignment Scores:

Pred. No.:	1.15e-121	Length:	368
Score:	1276.00 <td>Matches:</td> <td>251 </td>	Matches:	251
Percent Similarity:	80.17% <td>Conservative:</td> <td>36 </td>	Conservative:	36

Best Local Similarity: 70.11% Mismatches: 47  
 Query Match: 73.54% Indels: 24  
 DB: 2 Gaps: 3  
 US-10-617-962-2 (1-1008) x AAW97810 (1-368)

```

QY 4 GTTATACATTTAAACCTGATGATGA-----AGTGATATTCACCC 45
  |||||
DB 11 ValIleGlnLeuThrProAspAspArgValThrProAspAspLysGlyGluTyrGlnPro 30
  |||||
QY 46 GTTGAAGAACAAATAGCAGAGATATAGACGTATTAACCTTAAGGAAACAGATGAG 105
  |||||
DB 31 ValGluLysGlnIleAlaGlyAspIleIleArgValLeuGlnPheLysGlnThrAsnGln 50
  |||||
QY 106 GGTCTACAGCATCATGAAATGGAATTCAGACTAGAGAAATATATATTAAGCTTAACGT 165
  |||||
DB 51 SerHisThrGlyLeuTyrGlyIleAlaTyrArgAlaLysLysValIleIleAlaTyrAla 70
  |||||
QY 166 TTGGCTGTAGAGTATTTATATATGATTAACCTTCCGATGACCTATTTATAGAAATAA 225
  |||||
DB 71 LeuAlaValSerGlyIleHisAsnValSerGlnLeuProGluAspTyrTyrLysAsnLys 90
  |||||
QY 226 GAGACTGCTGAGAAATTTATCAAGAAATATATGCTTAATCTTCACTGACACTATTAAGT 285
  |||||
DB 91 AspaenthrGlyArgIleTyrGlnGluTyrMetSerHisLeuLeuSerAlaLeuLeuGly 110
  |||||
QY 286 GAAATATGATGATCAATTTCTAAAGATATGCAAAATGTTTATATAGAAATGAATGAT 345
  |||||
DB 111 GluAsnGlyAspGlnIleSerLysAspMetAlaAsnAspPheThrGlnAsnGlnLeuGln 130
  |||||
QY 346 TTTGAAGGTCAATATCCCTCAAAACATTTGGAAATGTTCCGAGCTTGAAATTAACCATG 405
  |||||
DB 131 PheGlyGlyGlnArgLeuLysAsnThrTyrAspIleProAspLeuGlnLysLeuLeu 150
  |||||
QY 406 AGTGCTTATTCAGATACGATTAATATATATGACCTATATTTTCTGTCAGACGAAAT 465
  |||||
DB 151 GluAspTyrSerAspGlnLysLeuLeuAlaLeuTyrPhePheAlaSerGlnGlnLeu 170
  |||||
QY 466 CCAGCTGAGAGAAATCAACAAATCAATGCGCGCAAGATTTTAAATTAATGATTTCTTA 525
  |||||
DB 171 PrometGluAlaAsnGlnGlnSerAsnAlaAlaAsnPheLysValIleAspPheLeu 190
  |||||
QY 526 TTTACCTTATTCGCTGTAACCTTCACTGAGAGAGAGATTTTTCAAAAACCTTTACAT 585
  |||||
DB 191 LeuIleLeuSerAlaValThrSerLeuGlyLysArgIlePheSerLysAsnPheTyrAsn 210
  |||||
QY 586 GGATTAAGAGGTAAATCATTTAGAGATTTATTTAGAGAAAAAACTTCTTAACCTTTC 645
  |||||
DB 211 GlyLeuGlnThrLysSerLeuGlnAsnTyrIleGluArgLysLysLeuSerLysProPhe 230
  |||||
QY 646 TTTGACCAACCGCAGAGATTAACCTGATGCGAGAAATAGTTATTTGGCTGACCAACAGAA 705
  |||||
DB 231 PheArgProProGlnLysLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrLys 250
  |||||
QY 706 GGGCCTAAA----- 714
  |||||
DB 251 AlaProLysLeuProThrThrSerSerThrAlaThrThrSerThrAlaAspSerSerAsn 270
  |||||
QY 715 TGGAGAGTGAATTTAAAGAACTTAAATAAACAATCTTAGAGATGATTTTCTTAATGAG 774
  |||||
DB 271 ThrPheValSerLeuGlnLysLeuArgAspAsnProSerArgAsnThrPheMetLysMet 290
  |||||
QY 775 GAAGGGGCTGCAAAAACAAAGATATAGTTCATTTATTAAGGATCAAAAAGGTTAAACGT 834
  |||||
DB 291 AspaAspAlaAlaLysArgLysTyrSerSerPheIleLysGluValGlnLysGlyAsnAsp 310
  |||||
QY 835 CCACAGACAGCAGCGAAAGATATGTTGTAACGACAGCGCATTAACCTGCAAAAATTTGCCG 894
  |||||
DB 311 ProArgAlaAlaAlaLysSerIleGlyThrLysSerGlySerAsnAsnHeGlnLysLeuGln 330
  |||||
QY 895 ---AATATTTTATATGATGAGGTAGGCTAGCAAAAAGACAGGCTTAACCTTAACAAAT 951
  |||||
DB 331 GlyArgAspLeuTyrSerIleArgLeuSerGlnGlnHisArgValThrPheSerIleAsn 350
  |||||

```

QY 952 GATCTGACAAATACATGACGTTGATGATGTTGGAACCTCATTTATTAATAATA 1005  
 ::|||  
 DB 351 AsnThrAspGlnIleMetGluIleGlnSerValGlyThrHisTyrGlnAsnIle 368

## RESULT 4

ID AAR68961 standard; protein; 278 AA.

XX AAR68961;

AC AAR68961;

DT 12-FEB-2004 (revised)

DT 27-AUG-2003 (revised)

DT 09-AUG-1995 (first entry)

XX Xenorhabdus nematophilus strain A24 insecticidal toxin.

XX Bacterium; insecticide; pesticide; toxin.

XX Xenorhabdus nematophila.

XX WO9500647-A1.

XX 05-JAN-1995.

XX 24-JUN-1994; 94WO-AU000348.

XX 25-JUN-1993; 93AU-00009638.

XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.

XX Smigajski AJ, Akhurst RJ;

XX WPI; 1995-052084/07.

XX N-PSDB; AAQ80839.

XX Polynucleotide(s) that encode insecticidal toxins from Xenorhabdus -

XX useful for the control of insect pests in the agricultural, aquatic and

XX forest industries.

XX Claim 8; Page 9; 15pp; English.

XX This insecticidal toxin which is useful for genetically engineering a

XX wide range of biological systems which will become more useful for the

XX control of insect pests detrimental to agricultural, aquatic and forest

XX industries. The protein may be expressed in recombinant organisms.

XX CC (updated on 25-MAR-2003 to correct PN field.) (updated on 27-AUG-2003 to

XX correct OS field.)

XX SQ Sequence 278 AA;

XX Alignment Scores:

XX Pred. No.: 2,666-87

XX Score: 940.50

XX Percent Similarity: 80.22%

XX Best Local Similarity: 71.27%

XX Query Match: 54.21%

XX DB: 2 Gaps: 2

US-10-617-962-2 (1-1008) x AAR68961 (1-278)

```

QY 4 GTTATACATTTAAACCTGATGATGA-----AGTGATATTCACCC 45
  |||||
DB 11 ValIleGlnLeuThrProAspAspArgValThrProAspAspLysGlyGluTyrGlnPro 30
  |||||
QY 46 GTTGAAGAACAAATAGCAGAGATATAGACGTATTAACCTTAAGGAAACAGATGAG 105
  |||||
DB 31 ValGluLysGlnIleAlaGlyAspIleIleArgValLeuGlnPheLysGlnThrAsnGln 50
  |||||
QY 106 GGTCTACAGCATCATGAAATGGAATTCAGACTAGAGAAATATATATTAAGCTTAACGT 165
  |||||
DB 51 SerHisThrGlyLeuTyrGlyIleAlaTyrArgAlaLysLysValIleIleAlaTyrAla 70
  |||||

```

```

QY 166 TTGGCTGTAAGTGTATTCATATGATGATTAACCTCCGATGACTATATTAAGATAAA 225
Db 71 LeuAlaValSerGlyIleHisValSerGlnLeuProGluAspTyrTyrIleAsnLys 90
QY 226 GAGACTGCTGAGAGAAATTTATCAGAAATATGCTTAATCTTTCATCTGCACTATTAAGT 285
Db 91 AspAenThrGlyArgIleTyrGlnValTyrMetSerAenLeuSerAlaLeuLeuGly 110
QY 286 GAAATGCTGATCAATTTCTTAAGATATGCAATGTTTATTAAGAAATGATGAT 345
Db 111 GluAsnGlyAspGlnIleSerIleAspMetAlaAsnAspPheThrGlnAsnGluLeuGlu 130
QY 346 TTGAGAGCTCAATATCTCAAAACATTTGATGATCTTCCGAGCTTGAATAATCAATG 405
Db 131 PheGluValAsnValLeuValIleProGlyIlePheLeuIleLeuAlaGlyIleAsnTyr 150
QY 406 AGTCCTTATTCAGATGACGATAAATTAATGACATATATTTTCTCTGACAGAAAT 464
Db 150 PylsIleTyrSerAspGluAspLysLeuLeuAlaLeuTyrPhePheAlaSerGlnGlu 170
QY 465 TCCACTGGAGGAAATCAACAATCAATGCGCGCAAGATTTTATTAATTAATGATGAT 524
Db 170 uProMetGluAlaAsnGlnIleSerAsnAlaAlaAsnPhePheLysValIleAspPhe 190
QY 525 ATTTACTTATCTGCTGTAATCTTCACTGCGAAGAGATTTTTCATAAAACCTTTACA 584
Db 190 uLeuIleuSerAlaValThrSerLeuGlyLysArgIlePheSerLysAsnPheTyrAs 210
QY 585 TGGATTAGAGCTTAATCATTAAGATTAATTAATGAGAGAAAAAATTTCTTAAACCTT 644
Db 210 nGlyLeuGlnThrLysSerLeuGluAsnTyrIleGluArgLysLysLeuSerLysProPh 230
QY 645 CTTTGACACCGCGGAGATTAACCTGATGCGCAATAGGTTATTTGGCTGACCAACAGA 704
Db 230 ePheArgProGlnLysLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrLys 250
QY 705 AGCGCTCAAA----- 714
Db 250 sAlaProLysLeuProThrThrSerSerThrAlaThrThrSerThrAlaAlaSerSerAs 270
QY 715 TGGAGAGTGAAGTTTAAAGAA 735
Db 270 nTyrArgValSerLeuGlnLys 277

```

RESULT 5  
AAP50777  
ID AAP50777 standard; protein; 1654 AA.  
XX  
AC AAP50777;  
XX  
DT 25-MAR-2003 (revised)  
DT 30-SEP-1991 (first entry)  
XX  
DE Sequence of the P195 protein of Plasmodium falciparum.  
XX  
KW Malaria vaccine; epitope; antigen; immunogen.  
XX  
OS Plasmodium falciparum.  
XX  
PN EP154454-A.  
XX  
PD 11-SEP-1985.  
XX  
PF 21-FEB-1985; 8SEP-00301173.  
XX  
PR 22-FEB-1984; 84GB-00004692.  
PR 26-SEP-1984; 84GB-00024340.  
PR 21-FEB-1985; 85GB-00004429.  
XX  
XX (WELL ) WELLCOME FOUND LTD.  
XX  
XX Holder A, Sandhu J, Odink K, Lockyer M, Riverosmor V;  
XX

```

DR WPI, 1985-224845/37.
DR N-PSDB; AAN50530.
XX
PT Cloned DNA sequence encoding plasmodium falciparum protein - useful for
PT expressing the protein for use in vaccines against malaria.
XX
PS Claim 6; Fig 1; 51pp; English.
XX
CC The sequence encoding the P195 protein of Plasmodium falciparum
CC (AAN50530) and a peptide comprising at least one of its epitopes (see
CC AAP50777) are claimed. Also claimed is a vaccine for inducing immunity to
CC malaria comprising the novel peptide or P195 or a peptide comprising at
CC least one epitope when derived from the new DNA sequence, together with a
CC carrier. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 25-MAR
CC -2003 to correct PI field.)
XX
SQ Sequence 1654 AA;

```

Alignment Scores:	Pred. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Score:	0.0934	1654	86	50	117	20
Percent Similarity:	108.50		86	50	117	20
Best Local Similarity:	34.96%		136	117		
Query Match:	6.25%					
DB:	1					

US-10-617-962-2 (1-1008) x AAP50777 (1-1654)

```

QY 121 TATGGAATGGAATATCGAGCT-----AGAAAAATATATATGACTTAC 162
Db 205 PheAsnLeuLysIleArgAlaAsnGluLeuAspValLeuLysLysLeuValPheGlyTyr 224
QY 163 GCTTGCTGTAAGTGTATTCATATGATGATTAACCTCCGAGCTTGAATATTAAG 219
Db 225 ArgLysProLeuAspAsnIleLysAspAsnValGlyLysMetGluAspTyrIleLysLys 244
QY 220 AATAAGAGACTGCTGAG----- 237
Db 245 AsnLysLysThrIleGluAsnIleAsnGluLeuIleGluGluSerLysLysThrIleAsp 264
QY 238 -----AGAATTATCAA--GAATATATG 258
Db 265 LysAsnLysAsnAlaThrLysGluGluGluLysLysLysLysLysLysLysLysLys 284
QY 259 TCTAATCTTCAATCGCACTATTAGTGAATAATGTCATCAATTTCT-----AAAGAT 312
Db 285 LeuSerIleTyrAsnLysGlnLeuGluGluAlaHisAsnLeuIleSerValLeuGluLys 304
QY 313 ATGCAATATGTTTATAGATGA-----CTGAT----- 345
Db 305 ArgIleAspThrLeuLysLysAsnGluAsnIleLysGluLeuLeuAspLysIleAsnGlu 324
QY 346 TTGAGAGCTCAATATCTCAAAACATTTGGAATGTTCTCTGAGCTT-----GAAAT 396
Db 325 IleLysAsnProProAlaAsnSerGlyAsnThrProAsnThrLeuLeuAspLysAsn 344
QY 397 AAACCATGAGTCTTATTCAGATGACATTAATTAATTAAGCACTATATTT-----TTCTCT 453
Db 345 LysLysIleGluGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 364
QY 454 GTACAGGAAT-----CCACTG----- 471
Db 365 IleAspSerLeuPheThrAspProLeuGluLeuGluTyrTyrLeuArgLysLysAsnLys 384
QY 472 -----GAGAAATCAACAATCA 489
Db 385 AsnIleAspIleSerAlaLysValGluThrLysGluSerThrGlnProAsnGluTyrPro 404
QY 490 AATGCCGCAAGATTT-----TTTAATTAATTAATTTCTTATTAATTAATTAATTAAT 546
Db 405 AsnGlyValThrTyrProLeuSerTyrAsnAspIleAsnAsnAlaLeuAsnGluLeuAsn 424
QY 547 TCACCTGGAGAGAGATTTT-----TCAAAAAAATTTTAC 582

```

```
Db 425 SerPheGlyAspLeuIleAsnProPheAspTyrThrLysGluProSerLysAsnIleTyr 444
Qy 583 AATGATTAGAGCTAAATCATTAAGAAATTAATTAAGAGAAA---AACTTTCTAAA 639
Db 445 ThrAspAnGluValArgLysPheIleAsnGluIleLysGluLysIleLysIleGluLys 464
Qy 640 CCTTTCTTTCAGCAACCCGACAGATTACTGATGCGAATAAGTTATTTGGCTGACCA 699
Db 465 LysLysIleGluSerAspLysSerTyrGluAspArgSerLysSerLeuAsnAspIle 484
Qy 700 ACAGAAAGCGCCTAATATGAGAGCTGATTAAAGAACTTAATAAATAAAGAACTAGAT 759
Db 485 ThrLys-----TTTCTAATATGAGAGCGCTGCAAAACAAAGATATGTTCAATTATAA 813
Qy 760 GGA-----TTTCTAATATGAGAGCGCTGCAAAACAAAGATATGTTCAATTATAA 813
Db 503 AsnIleAspLeuThrAsnPheGluLysMetMetGlyLysArgTyrSer-----Tyr 519
Qy 814 GAGGTACAAAGAGGTAAACGCTCCACAGACAGACGCGAAAGATTTGGTACAGCCAGTGC 873
Db 520 LysValGluLysLeuThrHisProAsnThrPheAlaSerTyr-----GluAsnSerLys 537
Qy 874 AGTAACTGGAATAAATTCGCGAATAATTAA-----TATGCTGAGAGGCTA 918
Db 538 HisAsnLeuGluLysLeuThrLysAlaLeuLysTyrMetGluAspTyrSerLeuArg--- 556
Qy 919 AGCCAAAAAGACAGGGTAACTTTACTCAAAAAGATACATGACATGACGGTTCAAT 978
Db 557 -----AsnIleValValGlu 561
Qy 979 AGTGTGAACTCATTTATAAATATA 1005
Db 562 LysGluLeuLysTyrTyrLysAsnLeu 570

RESULT 6
ABB47334
ID ABB47334 standard, protein, 940 AA.
XX
AC ABB47334;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #38.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR001118.
XX
PR 11-APR-2000; 2000FR-00004629.
XX
PA (INSP ) INST PASTEUR.
XX
PI Buchriesser C, Frangoul L, Couve E, Ruenick C, Feini H, Dehoux P,
PI Duesuget O, Chehouani F, Nedjati H, Glaeser F, Kunst F, Cossart P,
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L,
PI Perez-Diaz J, Baguero F, Garcia Del Portillo F, Gomez-Lopez N,
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Haut J,
PI Rose M, Voss H;
XX
DR WPI, 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
and prevention of Listeria and related bacterial infections, and related
```

```
PT polypeptides.
XX
PS Claim 6, SEQ ID NO 39, 192pp; French.
XX
CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms,
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 940 AA;
XX
Alignment Scores:
Pred. No.: 0.0885 Length: 940
Score: 108.00 Matches: 75
Percent Similarity: 37.03% Conservative: 52
Best Local Similarity: 21.87% Mismatches: 136
Query Match: 6.22% Indels: 80
DB: 5 Gaps: 16

US-10-617-962-2 (1-1008) x ABB47334 (1-940)
Qy 112 ACAGATCATATGAAATGAAATATGAGCT-----AAGAAATAATATTAAGCTTAC 162
Db 355 ThrValThrTyrLysAlaLysPheThrAlaThrSerLysGlyValProLeuSerTyr 374
Qy 163 GCTTGGCTGTAAAGTGTATTCATATGATTAATCTAACTTCCTGATAC----- 210
Db 375 SerIleAsnValSerGlnProIleAsnValSerGluGlnThrAspSerThrValSerVal 394
Qy 211 TATTTAAAGATTAAGACAGCTGCTGAGAAATTTATCAAGAAATATATGCTTAATCTTCA 270
Db 395 PheTyrGlnAspGluAsnGlyAsnGluLeuAlaProThrGluThrLeuSerGlyLysSer 414
Qy 271 TCTGCATATTAGCGTGAATAATGCTGATCAAAATTTCTAAGAAATATGCAAAATGTTTAT 330
Db 415 -----GlyGluAspTyrGlnThrThrGluLysThrIleAlaAsn----- 427
Qy 331 AAGATGAACGATGATTTGAAGTCAATATCTCAAAACATTTGGATGTTCTGAGCTT 390
Db 428 -----TyrGlnLeuLysGluIle 433
Qy 391 GAAATTAACCATGAGTCTTATTCAGATGACGATAAATATTATGACATATATTT--- 447
Db 434 GluGlyGlnAlaSerGlyGlnPheThrAspThrAspSerThrValThrTyrValTyrGlu 453
Qy 448 -----TTCTGTGACAGAAATTCACACGACGAGAGAAATGACAAATCA 489
Db 454 LysAlaAspGlyAlaProValThrValLysTyrValAlaAspGlyAsnAspLeuAla 473
Qy 490 AATGCCGCAAGATTTTAAATTAATTAATGATTTCTTATTAATCTTTCGCTGTAATCTCA 549
Db 474 ThrSerAspThrLeuAsnGlyLysIleAspAlaProTyrGlnThrSerAlaLysSerLeu 493
Qy 550 CTGGAAGAGGATT---TTTCAAAAACCTTTTCAATGAGATT-----GAGCTAAA 600
Db 494 SerGlyTyrThrValLysThrThrProAsnAsnAlaThrGlyValPheThrAsnSerLys 513
Qy 601 TCATTATGAGAAATTAATTAATGAGAAATAAATTTCTTAATCCTTTTTCGACACCGGAG 660
```





```

Db      448 ProGluValIleThrGlnThrAsnProArgGlnAsnIleIleThrLysAsnAlaAsnHis 467
Qy      724 -----AGTTTAAAGAACTTAAATAACAAATATGAGATGATTTCTAAT 771
Db      468 IleArgValAlaSerPheAsnValLeuAsnTyrAspAsnGlyAlaThrGlyPheProThr 487
Qy      772 ATGGAAGGGCGCTGCAAAACA-----AAGTAAATGTTCAATT 807
Db      488 GluArgGlyAlaAsnThrGlnAlaGluPheAspLysGlnHisIleLysIleValSerAla 507
Qy      808 ATAAAGAGGTA-----CAAAAGGTAAACGT 834
Db      508 LeuLysSerIleAspAlaAspValTyrGlyLeuMetGluIleAlaAsnAsnGlyTyrGly 527
Qy      835 CCACAGACGACGCAAAAGTATGGTACAGCCAGCTGC----- 873
Db      528 ProAsnSerAlaIleAlaHisIleThrSerAlaLeuGlyProAspTyrLysTyrValIle 547
Qy      874 ---AGTAACTGGAAAATTTGCCGAATTAATTATATAGTGAAG----- 915
Db      548 ProGluAsnLeuAspArgLeuGlyAsnAspValIleAlaValAlaIleTyrAsnSer 567
Qy      916 -----CTAAGCCAAAAGACAGGGTA 936
Db      568 LysArgValLysProLeuAsnLysAlaValValLeuAspLeuGlyAspLysAsnArgThr 587
Qy      937 ACCTTACTCA----- 948
Db      588 ThrLeuAlaGlnThrPheGlnAlaValArgGlyAsnLysIlePheThrValIleProAsn 607
Qy      949 -----AATGATACGACCAATACA 966
Db      608 HisLeuLysSerLysGlyCysSerGlyValAspAlaSerSerSerAspAlaAspGln--A 627
Qy      967 ATGAGGTTTCATAGTCTTGGAAAC 989
Db      627 snAspGlyGlnGlyCysTrpAsn 634

RESULT 8
ADE83729
ID      ADE83729 standard; protein; 1270 AA.
XX
AC      ADE83729;
XX
DT      29-JAN-2004 (first entry)
XX
DE      Bacillus thuringiensis cry6 insecticidal protein #1.
XX
KM      insecticidal protein; TK-E6; FERM P-17981; cry6; insecticide.
XX
OS      Bacillus thuringiensis.
XX
PN      JP200235967-A.
XX
PD      26-NOV-2002.
XX
PF      16-MAY-2001; 2001JP-00146729.
XX
PR      16-MAY-2001; 2001JP-00146729.
XX
PA      (UYKI-) UNIV KINKI.
XX
DR      MPI; 2003-485827/46.
XX
DR      N-PSDB; ADE83728.
XX
PT      New Bacillus thuringiensis strain (Bacillus thuringiensis TK-E6 (FERM P-17981)) producing an insecticidal protein.
XX
PS      Claim 3; SEQ ID NO 2; 29pp; Japanese.
XX
CC      The invention comprises a microorganism that is capable of producing an insecticidal protein - Bacillus thuringiensis TK-E6 (FERM P-17981). The

```

```

CC      invention also comprises the amino acid and coding sequences of a number
CC      of insecticidal proteins (cry6 proteins). The bacterium and cry6
CC      proteins of the invention are useful as insecticides. The present amino
CC      acid sequence represents a Bacillus thuringiensis cry6 protein of the
CC      invention.
XX
SQ      Sequence 1270 AA;
XX
Alignment Scores:
Pred. No.:      0.175      Length:      1270
Score:          105.50      Matches:      69
Percent Similarity: 34.19%      Conservative: 38
Best Local Similarity: 22.04%      Mismatches: 119
Query Match:      6.08%      Indels:      87
DB:              7      Gaps:      14

US-10-617-962-2 (1-1008) x ADE83729 (1-1270)
Qy      13 TTAACACCTGATGATAGAAAGTGATATCCACCGTTGAAAAG----- 54
Db      544 LeuIleProAlaAsnLysIleGlyGlnIleProAlaValLysGlyAsnTyrLeuGlyGln 563
Qy      55 -----CAATAGCAGAGATATGATGATACTA 84
Db      564 SerGlySerSerValValArgThrSerGlyAsnThrGlyGlyAspValValArgLeuTyr 583
Qy      85 AACTTAAAGCAAAACAGATGAGGGTCAACAGCATCATATGGAATTAATGACCTAAG 144
Db      584 GluPheGlyAsnLeu-----GlyMetThrValArgPheSer--GluAsnArgSerTyr 600
Qy      145 AAAATTAATTAATTAAGCTTACGCTTGGCTGTAAGTGATATCATATGTA----- 192
Db      601 IleIleArgLeuArgTyrAlaThrValAlaAspAspLeuAsnIleIleValArgValGln 620
Qy      193 -----TCTAACTTCTGATGACTATTATAGATATAAGAG 228
Db      621 ArgArgGlyGluLeuGlnTyrGluSerGluLeuProLeuAsnGlnThrSerAsnAsnSer 640
Qy      229 ACTGCTGAGAGATTTATCAAGATATATGTTCTAATCTTCATCTGCACTATATAGGTGA 288
Db      641 ThrThrGlnTyrPlyPheGluAspTyr----- 649
Qy      289 AATGCTGATCAAAATTTCTAAAGATATGCAAAATGCTTTTATAGAAATGACGATTTT 348
Db      650 -----GlyTyrGln-----GluVal 654
Qy      349 GAAGGCAATATCCCTCAA-----AACATTGGGAATGCTCGAGCTTGA 393
Db      655 GlyIlePheTyrProGlnValGlyGluGluTyrGluLeuTyrPheSerPro-----Val 672
Qy      394 AATAAACCATTCAGTGTATTCAGATGACGATTAATTAATTAAGCACTATATTTTCTCT 453
Db      673 GlyThrGluLeuSerSerHisMetAspIleAspLysIle-----GluPheIle 688
Qy      454 GTACAGAAATTCACCTGAGAGAA-----AATCAACATCAAAATGCCCAAGATTT 504
Db      689 ProMetGluThrSerLeuGluGluTyrLeuAlaAsnGlnAspIleGluLysAlaArg--- 707
Qy      505 TTTAATTAATTAATTAATTTTATTTACCTTATCTGCTGTAACCTGCGAAGAGGATTT 564
Db      708 ---LysAlaValAsnAlaLeuPheThrGlyAspValLysAsnAlaLeuLysLeuAsnVal 726
Qy      565 TTTTCA-----AAAACTTTTACATGATGATTAAGGCTTAATCACTTAAGATTAAT 615
Db      727 ThrAspTyrAlaIleAspGlnAlaAlaAsnLeuValGluCysValSerGluGluPheTyr 746
Qy      616 ATTGAGAGAAAAAATCTTCTTAACCTTTCTTTCAGACCAACCGCAGAGATTAACCTATGCGC 675
Db      747 AlaGlnGluLysMetIleLeuLeuAspGlnValLysValAlaLysArgLeuSerGlnAla 766
Qy      676 AGAATAGTTATTTGGCTGACCAACAGACGCGCTTAATG----- 717
Db      767 ArgAsnLeuLeuAsnTyrGlyAspPheGluSerProGluTyrPheSerArgGluAsnGlyTyr 786

```

QY 718 -----AGAGTGGTTTAAAGACTTAAATACAAATCTAGAGATGATTT 765  
 DB 781 LysThrSerArgHisValSerValArg---AlaAspAsnProValPheLysGlyArgTyr 805  
 QY 766 TCTAATATGGAAGGGCTGCAAAACAAAGTATAGTTCA 804  
 DB 806 LeuHisMetProGlyValThrSerProSerPheSerAsn 818  
 RESULT 9  
 ADK16573  
 ID ADK16573 standard; protein; 2197 AA.  
 AC ADK16573;  
 DE 06-MAY-2004 (first entry)  
 XX Nanoarchaeum equitans cancer-associated (CA) protein #262.  
 XX Nanoarchaeum equitans cancer-associated (CA) protein; lymphoma; leukaemia.  
 OS Nanoarchaeum equitans.  
 XX WO2003093434-A2.  
 XX 13-NOV-2003.  
 PF 01-MAY-2003; 2003WO-US013699.  
 XX 01-MAY-2002; 2002US-0377447P.  
 PA (DIVE-) DIVERSA CORP.  
 PI Stetter KO, Waters E, Kretz K, Podar M, Richardson T,  
 PI Noordeweyer M;  
 DR WPI: 2004-053041/05.  
 DR N-PSDB; ADK16572.  
 XX New recombinant cancer-associated genes, such as KCNU9, useful for  
 PT diagnosing or treating carcinoma, e.g. breast, colon, rectal, pancreatic,  
 PT cervical, or skin cancers, lymphomas, or leukemia.  
 XX Claim 64; SEQ ID NO 525; 251pp; English.  
 CC The invention comprises then amino acid and coding sequences of cancer-  
 CC associated (CA) genes isolated from Nanoarchaeum equitans. The invention  
 CC also comprises the Nanoarchaeum equitans genome. The DNA and protein  
 CC sequences of the invention are useful for diagnosing and treating cancer  
 CC (e.g. carcinoma, lymphoma, or leukemia). The present amino acid sequence  
 CC represents a Nanoarchaeum equitans CA protein of the invention.  
 XX  
 SQ Sequence 2197 AA;  
 Alignment Scores:  
 Pred. No.: 0.262 Length: 2197  
 Score: 104.50 Matches: 72  
 Percent Similarity: 38.07% Conservative: 54  
 Best Local Similarity: 21.75% Mismatches: 154  
 Query Match: 6.02% Indels: 51  
 DB: 8 Gaps: 13  
 US-10-617-962-2 (1-1008) x ADK16573 (1-2197)  
 QY 73 GTAGCTATACCTTAAGCAACAGATGAGGCTCATACAGATCATATGGAATTGAA 132  
 DB 670 ValLeuLeuSerAsnIleArgAsnAsnSerTyrAlaThrLeuSerAsnTyrLeuIleHis 689  
 QY 133 TATCGAGCTTAAGAAATATATATAGCTTACGCTTGAGCTGTAAGTGTATCATATGTA 192  
 DB 690 TyrAlaAlaTyrIleValValAsnAspIleValAsnGluIleIleAsnGluLeuAsnIle 709  
 QY 193 TCTAAACTTCCTGATGACTAT-----TATTAAGATATAAGAGACTGCTGAGAAATTAT 246

DB 710 -----AspAspPheValGlnTyrAsnAsnLys-----LysGluLeuIle 722  
 QY 247 CAAGATATATAGCTTAATCTTTCATCTGCACATATAGCTGAAATAGTGATCAATTTCT 306  
 DB 723 LysAsnTyrPValSerAsnSerIleSerHisCysLeuGlyGlu----- 736  
 QY 307 AAAGATATGCAAAATGTTTATTAAGAATGATGATGATTTTGAAGTCAATATCTGCA 366  
 DB 737 ValAspIleLeuMet---PheLeuThrAlaAsnIleAspLeuGluArgLeuTyrGln 755  
 QY 367 AACATTTGAAATGTTCTGACCTTGA---AATAAACATTTAGTCTTATTCAGATGAC 423  
 DB 756 GluLeuGlnAsnTyrIleAlaLeuSerValLeuThrProLeuSerIleTyrSerSerAsn 775  
 QY 424 GATTAATTTATGACCTATATTTTCTCTGTACAGAAATTCCTGAGAGAAATATCA 483  
 DB 776 Glu-----ValIleTyrIleArgGluMetValLysSerGlnLeuGluMetGlyLeu 792  
 QY 484 CAATCA-----AATGCCGCAAGATTTTAAATTAAT 516  
 DB 793 SerSerLeuAspAsnLysLeuGluProIleIleAsnThrLeuThrAsnSerLeuTyr 812  
 QY 517 GATTTCTTATTTACCTTATCTGCTGTAACCTTCACTGGAAGAGGATTTTCAAAAAAC 576  
 DB 813 AsnLeuAlaTyrSerLeuSerGlnValThrSerSerLeuThrAsnTyrPheSer---Ser 831  
 QY 577 TTTTCAATGATTTAGAGCTTAATCATTTAGAGATTTATTTAGAGAAAAAACTT--- 633  
 DB 832 ValValSerSerIleAsnSerIleValLeuProSerThrIleGlyGluIleLysLeuIle 851  
 QY 634 -----TCTAAACCTTTCTTGACACGACGAGATTCCTGATGCGAGATA 681  
 DB 852 AspTyrLysThrLeuLysAlaIleAlaLeuIleProSerSerValGlnAlaIleIle 871  
 QY 682 GGTATTTTGCTGACCAAGAGAGCGCTTAATAGAGAGTGTATTAAGAACTTAA 741  
 DB 872 AsnArgLeu-----ThrAsnLysValSerTyrAsnIleLeuLeuGluCysIleAsp 888  
 QY 742 AATAACAAATCTAGAAATGATTTCTTAATATGCAAGGGCTGCAAAACAAAGTATAGT 801  
 DB 889 ThrLysLeuSerLysAlaAlaAsnIleGlnIleAlaLeuAlaLysLeuGlyTyrLeu 908  
 QY 802 TCATTATTAAGAGTCAAAAGGTAACGCTCACAGACAGCGCAAAAGTATGAT 861  
 DB 909 SerIleIleGluGlyIleAspProAspSerTyrProAspLysLysAlaGln----- 925  
 QY 862 ACAGCAGTGGCAGTACCTGGAATAATTCGGAATTAATTATATAGTGTAGGCTTAAC 921  
 DB 926 -----LeuGluGlnLeuIleSerLysValLysGluValLysLysIle 939  
 QY 922 CAAAAACAGGGTACCTTACTCAAAATGATCTGACATACATACATACGTTCATAGT 981  
 DB 940 IleGluGlnGlnGlnSerThrLeuLysLysAspLeuAspAsnThrIleAsnLeuTyrLys 959  
 QY 982 GTT-----GGAACCTATTAATAAAT 1002  
 DB 960 GluPheIleLysGlySerThrGluTyrIleAsn 970  
 RESULT 10  
 ABR63439  
 ID ABR63439 standard; protein; 583 AA.  
 AC ABR63439;  
 DE 08-SEP-2003 (first entry)  
 XX S suis aro gene cluster encoded protein #1.  
 DE Avirulent vaccine; antibiotic; aro gene cluster; immunisation;  
 KW antibacterial; gene therapy.  
 OS Streptococcus suis.







CC which is attached thereto from cells. Such protein complexes may comprise  
CC up to 30 distinct proteins. Protein complexes of the invention are useful  
CC for diagnosing a disease or disorder, or as a target for an active agent  
CC of a pharmaceutical, preferably a drug target in the treatment or  
CC prevention of a disease or disorder. Note: The sequence data for this  
CC patent is not represented in the printed specification, but is based on  
CC sequence information supplied by the European Patent Office. The complete  
CC document is available on CD-ROM

XX  
XX Sequence 566 AA:

Alignment Scores:  
Pred. No.: 0.396 Length: 566  
Score: 101.00 Matches: 46  
Percent Similarity: 38.01% Conservative: 38  
Best Local Similarity: 20.81% Mismatches: 87  
Query Match: 5.82% Indels: 50  
Gaps: 8

US-10-617-962-2 (1-1008) x ABR53129 (1-566)

QY 442 TATTTTCTCTGTACAGAAATTCCTGAGGAAATCAAAATCCGCAAGA 501  
DB 27 PhePheHsAlaLylLysLysAspProValaEngInAspLysAlaAsnHsAlaSerGln 46  
QY 502 -----TTTTT 507  
DB 47 IleHrProThrValProHisSerHisProSerAspMetValIleProAspHisLeuAla 66  
QY 508 AAATTAATGATTTCTTATTTACTTATCTGCTGTAATCTTCACTGGAAAGAGATTTT 567  
DB 67 GluLeuIleProGluLeuLysSerPheGlnGlnLeuValAspSerGluLysArgLeuAsp 86  
QY 568 TCAAAAACCTTTACATGATGATTAGAGCTAAATCATTAAGAAATTAATTGAGAGAAA 627  
DB 87 HisPheIleHisLeuLysAsnLeuHisMetLysArgMetValAlaGlnTrpGluArgSer 106  
QY 688 AAACCTTTCAACCTTTCTTCTTGACACCGCAGAGATTTACCTGATGGCGAGATTA----- 681  
DB 107 LysLeuSerGlnGluPheLeuLysProHisLeuAsnPheProAsnValIlePheLeuArg 126  
QY 682 GGTATTTGGCTGACCAACAGAGAGCGCTTAATGAGAGATTTTAAAGAACTTAA 741  
DB 127 IlePheIleSerAsnValSerGlnAsnGlnProTrpGlnMet-----AspThrAsn 143  
QY 742 AATAACAATCTAGAAATGATTTCTTAAT-----ATGAGAGCG--- 780  
DB 144 AsnGluAlaAspLeuMetAlaLeuGlnAsnAlaThrTrpThrMetArgIleGluArg 163  
QY 781 -----GCTGAAAACAAAGATTAATGATTTCTTAAT 813  
DB 164 LeuLeuAspAsnValGlnAlaAsnAspProAlaAsnGlnLysLysPheSerPheIleGln 183  
QY 814 GAGGTA-----CAAAAGGTAACGCTCCACAGACA----- 843  
DB 164 SerIleValValAspPheLysAsnLysGlnAsnAspAsnValProSerThrLysPheAsn 203  
QY 844 GCAAGCAAAAGTATTGGTACAGCCAGTGCAGTAACCTGCAAAATTCGCCAATTAATTA 903  
DB 204 AlaAlaProGluGlnAsnAlaThrGlnGlyProSerAspLysLysLeuAsnLeuAsnLeu 223  
QY 904 ---TATAGTGTAGGCTAAGCCAAAAGACAGGTTAACTTTATCAAAATGATCTGAC 960  
DB 224 ProLeuGlnPheSerLeuProAsnGlnLysAsnSerThrThrThrAsnThrAspGlnAsn 243  
QY 961 AAT 963  
DB 244 Asn 244  
RESULT 14  
ADK62560  
ID ADK62560 standard; protein; 566 AA.  
XX

AC ADK62560;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX Disease treating protein complex-derived protein #406.  
XX  
XX protein complex; drug target; diagnosis.  
XX  
XX Unidentified.  
XX  
XX EP138608-A2.  
XX  
XX 27-AUG-2003.  
XX  
XX 20-DEC-2002; 2002EP-00102902.  
XX  
XX 20-DEC-2001; 2001EP-00130253.  
XX  
XX (CELL-) CELLZOME AG.  
XX  
XX Bauer A, Gavin A, Superfi-Furga G, Kuester B, Schultz J;  
XX  
XX Marzloch M, Grandi P, Krause R, Krause U, Merino A, Bauch A;  
XX  
XX Michon A, Leutwein C, Rick J;  
XX  
XX WPI, 2003-638460/61.  
XX  
XX N-PSDB; ADK62561.  
XX  
XX  
XX New proteins and protein complexes from eukaryotes, useful as targets in  
XX  
XX drug screening, or in diagnosing or screening for the presence of a  
XX  
XX disease or disorder, or a predisposition for developing a disease or  
XX  
XX disorder in a subject.  
XX  
XX Disclosure; SEQ ID NO 811; 13pp; English.  
XX  
XX  
XX The invention relates to novel protein complexes comprising a first and a  
XX  
XX second protein, or its derivative, fragment, homologue or variant. The  
XX  
XX proteins are selected from given protein complexes, which are not defined  
XX  
XX in the specification. The variants are encoded by nucleic acids that  
XX  
XX hybridize to the nucleic acids encoding the proteins under low stringency  
XX  
XX conditions. The protein complexes are useful as targets for an active  
XX  
XX agent of a pharmaceutical. These protein complexes are particularly  
XX  
XX useful as drugs targets for the treatment or preventing of a disease or  
XX  
XX disorder. The complexes and methods above are useful in diagnosing or  
XX  
XX screening for the presence of a disease or disorder or a predisposition  
XX  
XX for developing a disease or disorder in a subject. These are also useful  
XX  
XX in screening for a drug for treatment or prevention of a disease or  
XX  
XX disorder. The molecule that modulates the amount, activity or protein  
XX  
XX components of the complex is useful for the manufacture of a medicament  
XX  
XX for the treatment or prevention of a disease or disorder. This sequence  
XX  
XX corresponds to a protein of the invention. (Note: the sequence data for  
XX  
XX this patent did not form part of the printed specification but was  
XX  
XX obtained from the EPO in electronic format).

XX  
XX Sequence 566 AA:

Alignment Scores:  
Pred. No.: 0.396 Length: 566  
Score: 101.00 Matches: 46  
Percent Similarity: 38.01% Conservative: 38  
Best Local Similarity: 20.81% Mismatches: 87  
Query Match: 5.82% Indels: 50  
Gaps: 8

US-10-617-962-2 (1-1008) x ADK62560 (1-566)

QY 442 TATTTTCTCTGTACAGAAATTCCTGAGGAAATCAAAATCCGCAAGA 501  
DB 27 PhePheHsAlaLylLysLysAspProValaEngInAspLysAlaAsnHsAlaSerGln 46  
QY 502 -----TTTTT 507  
DB 47 IleHrProThrValProHisSerHisProSerAspMetValIleProAspHisLeuAla 66

```
QY 508 AATTAATGATTTCTTATTATCTTATGCTGTAATCTACTGGGAAGAGATTTT 567
Db 67 GtuleuileProgluileuYrSerPheglnglnleuValAspserGluYrAsnleu 86
QY 568 TCAAAAACCTTTACATGATTAAGAGGCTTAATCACTTACAGATTAATATTGAGAA 627
Db 87 HisPheleHleuYrAsnleuHISMetLysAsnMetValAlaGlnTyrGluYrSer 106
QY 628 AAACCTTTTAAACCTTTCTTTGACCAACCGACAGATTAATCTGATGGCAGAAATA----- 681
Db 107 LysleuSerGlnGluPheleuYrProHISleuAsnPheProAsnValLysPheleuYr 126
QY 682 GGTATTATGGCTGACCAACAGAGCGCTTAATGAGAGTGATTTTAAAGAACTTAA 741
Db 127 IlePheleSerAsnValSerGluAsnGlnProTyrGlnMet-----AspThrAsn 143
QY 742 AATAACAATCTAGAAATGATTTTCTAT-----ATGGAAGG----- 780
Db 144 AsnGluAlaAspLeuMetAlaLeuGluAsnAlaThrTyrThrMetArgIleGluYrArg 163
QY 781 -----GCTGCAAAACAAAGATATGTTCAATTAATAAA 813
Db 164 LeuLeuAspAsnValGlnAlaAsnAspProAlaArgGluLysPheSerSerPheIleGlu 183
QY 814 GAGGTA-----CAAAAGGTAACGCTCCACAGCA----- 843
Db 184 SerIleValAlaAspPheLysAsnYrGluAsnAspAsnValProSerThrLysPheAsn 203
QY 844 GCAGCGAAAGATTTGTTACGACGACGAGTACGTAACCTGTAATAATGCGAATATTA 903
Db 204 AlalaAlaProGluGluAsnAlaThrGluGlyProSerAspLysLysLeuAsnleu 223
QY 904 ---TATAGTGTAGGCTTAAGCCAAAAGAAGAGGTTACCTTACTCAAAATGATACTGAC 960
Db 224 ProLeuGlnPheSerLeuProAsnGlyAspAsnSerThrThrAsnThrAspGlnAsn 243
QY 961 AAT 963
Db 244 Asn 244

RESULT 15
ADP25426
ID ADP25426 standard; protein; 2404 AA.
XX
AC ADP25426;
XX
DT 09-SEP-2004 (first entry)
XX
DE Plasmodium falciparum antigen amino acid sequence SEQ ID NO:3.
XX
KW Plasmodium falciparum; malaria parasite; antigen; malaria; immunogenic;
KW immune response; cytostatic; anti-HIV; virucide; hepatotropic;
KW antibacterial; vaccine; cancer; infectious disease; AIDS; hepatitis;
KW bacterial infection.
XX
OS Plasmodium falciparum.
XX
PN WO2004053086-A2.
XX
PD 24-JUN-2004.
XX
PF 08-DEC-2003; 2003WO-US038966.
XX
PR 06-DEC-2002; 2002US-0431494P.
XX
PA (EPIM-) EPIMUNE INC.
XX (USNA ) US SEC OF NAVY.
XX
PI Sette A, Doolan DL, Carnuci DJ, Sidney J, Southwood S;
XX
DR WPI; 2004-468856/44.
XX
PT New isolated and/or purified Plasmodium falciparum polynucleotide
```

```
PT sequences, useful in inducing an immune response for preventing and/or
PT treating cancer and infectious diseases, such as AIDS, hepatitis, and
PT bacterial infections.
XX
PS Claim 22; SEQ ID NO 3; 253pp; English.
XX
CC The present invention describes an isolated and/or purified Plasmodium
CC falciparum (malaria parasite) antigen polynucleotide sequence, encoding
CC an immunogenic peptide. Also described: (1) a primer or detection probe
CC for hybridisation with a target sequence or the amplicon generated from a
CC target sequence comprising a sequence of at least 8-30, 35, 40, 45, 50,
CC 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 consecutive nucleotides of any
CC of the polynucleotide sequences as described above; (2) a DNA chip
CC comprising any of the polynucleotide sequences described above; (3) a
CC vector comprising a promoter operably linked to any of the nucleic acid
CC sequences described above; (4) a host cell transformed by the vector of
CC (3) or the polynucleotide described above; (5) a composition comprising a
CC carrier and the polynucleotide described above; (6) a method of inducing
CC an immune response in an individual comprising the administration of the
CC composition of (5) to induce an immune response; (7) an isolated
CC polypeptide comprising any of the amino acid sequences as encoded by the
CC polynucleotide described above; (8) a composition comprising a carrier
CC and the polypeptide of (7); (9) a method of detecting P. falciparum in
CC biological samples, comprising contacting a biological sample with the
CC isolated polynucleotide and detecting the hybridisation of the isolated
CC polynucleotides with nucleic acids contained in the sample; (10) a method
CC for eliciting an immune response in an individual, comprising the
CC administration of a composition comprising the polypeptides of (7) to an
CC individual to induce an immune response in the individual; (11) an
CC antibody that specifically binds to the P. falciparum polypeptide of (7);
CC and (12) detecting P. falciparum antigens, comprising contacting a sample
CC from a subject with the polypeptide of (7) and detecting the presence of
CC an antigen-antibody complex or detecting the stimulation of T-cells in
CC the sample. The P. falciparum antigens and immunogenic peptides have
CC cytosolic, anti-HIV, virucide, hepatotropic and antibacterial
CC activities, and can be used in vaccines. The methods and compositions of
CC the present invention are useful for inducing an immune response for the
CC prevention and/or treatment of cancer and infectious diseases, such as
CC AIDS, hepatitis, and bacterial infections. The present sequence
CC represents a P. falciparum antigen amino acid sequence, which is used in
CC the exemplification of the present invention.
XX
SQ Sequence 2404 AA;
XX
Alignment Scores:
Pred. No.: 0.779 Length: 2404
Score: 100.00 Matches: 49
Percent Similarity: 41.95% Conservative: 37
Best Local Similarity: 23.90% Mismatches: 54
Query Match: 5.76% Indels: 65
DB: 8 Gaps: 10

US-10-617-962-2 (1-1008) x ADP25426 (1-2404)
QY 142 AAGAAATTAATTAATGCTTACGCTTGGCTGTAATGCTATTCAT----- 186
Db 326 LysAsnThrIleAsnAlaPheSerIleAsnYrThrGlyValAsnPheGluAlaMetLys 345
QY 187 -----AATGATCTAAACTTCCTGATGAC-----TTTTTAAGATAAGAGACT 231
Db 346 GlnleuAsnAspLysAlaSerleuLeuPheAsnValYrYrGluLysLysGluAsn 365
QY 232 GCTGAGAGAAATTTATCAAGAAATATATGCTATATCTTATGCACTATTAGTGGAAT 291
Db 366 SerAsnArg-----GluGluIle 371
QY 292 GGTGATCAAAATTTCTAA-----GATATGGCAAAATGCTTTT 327
Db 372 AsnAspLysValSerLysGlnGlyCyAsnleuAsnAspSerAspSerSerAsnValleu 391
QY 328 TAT-----AAGATGAAGTGAATTTTGAAGTCAATATCCCAAAACATTTGGAAT 378
Db 392 TyrIleAsnIleGlnAsnIleLysAspTyrAspIleLeuYrLysGluAsnLysAsn 411
```





THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	109.5	6.3	414	2	H70219	hypothetical prote
2	108.5	6.3	1640	2	A24554	probable major sur
3	108	6.2	940	2	AD1374	internal protein
4	108	6.2	1127	2	T28317	ORF MSV156 hypothe
5	106	6.1	1785	2	A45546	major merozoite sur
6	104.5	6.0	755	2	T41912	structural phospho
7	104	6.0	1428	2	C85079	hypothetical prote
8	104	6.0	2708	2	T09079	probable chloroqu
9	103	5.9	579	2	H97194	uncharacterized AB
10	103	5.9	853	3	S60178	gag polyprotein h
11	103	5.9	950	2	T09076	hypothetical protei
12	102.5	5.9	2510	2	T28160	hypothetical prote
13	102	5.9	1132	2	T31107	telomerase reverse
14	101.5	5.9	657	2	AD1575	probable cell sur

15	101.5	5.9	1539	2	S05603	major merozoitc su
16	101	5.8	537	2	G64432	hypothetical prote
17	101	5.8	566	2	L19063	hypothetical prote
18	100.5	5.8	1531	1	SA2QK1	major merozoitc su
19	100	5.8	978	2	A70387	conserved hypotc
20	99	5.7	679	2	S06000	penicillin-binding
21	98.5	5.7	616	2	G82885	hypothetical prote
22	98.5	5.7	634	2	D71493	probable DNA helic
23	98.5	5.7	834	2	B82940	conserved hypotc
24	98.5	5.7	1587	2	AB9012	hypothetical prote
25	98.5	5.7	1587	2	T18485	hypothetical prote
26	98	5.6	580	2	C81352	lipid export ABC t
27	98	5.6	745	2	H85048	probable transposo
28	98	5.6	954	2	T09750	exciNuclease ABC c
29	98	5.6	2819	2	T09080	probable chloroquin
30	97.5	5.6	546	2	T17665	probable t-complex
31	97	5.6	348	2	B69790	hypothetical prote
32	97	5.6	645	2	C84999	ATP-dependent DNA
33	97	5.6	780	2	T21708	hypothetical prote
34	97	5.6	1714	2	E81469	Sex/Thr protein Ki
35	96.5	5.6	252	2	H84423	hypothetical prote
36	96.5	5.6	680	2	G95194	penicillin-binding p
37	96.5	5.6	685	2	D98061	penicillinbinding p
38	96.5	5.6	1017	2	D90550	vea1-like (mycopla
39	96.5	5.6	2401	2	T28676	thoxy protein -
40	96	5.5	617	2	H90551	lipoprotein [ampor
41	96	5.5	623	3	S67762	hypothetical prote
42	96	5.5	8745	2	F81453	DNA-directed DNA p
43	96	5.5	3161	2	T03042	protein HMMP1 - Ye
44	96	5.5	3163	2	AB0233	Yersiniabactin bio
45	96	5.5	3163	2	T17440	probable polyketid

## ALIGNMENTS

```

RESULT 1
H70219
hypothetical protein BAB28 - Lyme disease spirochete plasmid B/cp26
C/Species: Borrelia burgdorferi (Lyme disease spirochete)
C/Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C/Accession: H70219
R/Author: C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kierlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A/Authors: Smith, H.O.; Venter, J.C.
A/Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A/Reference number: A70100; MUID:98065943; PMID:9403665
A/Accession: H70219
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-414 <KLB>
A/Cross-references: UNIRROT:O50999; GB:AE000792; NID:G3253098; PIDN:AAAC6335.1; PID:G3268
A/Experimental source: strain B31
C/Genetics:
A/Genome: plasmid

Alignment Scores:
Pred. No.:          1.2          Length:          414
Score:              109.50       Matches:          72
Percent Similarity: 36.02%      Conservative:     44
Best local Similarity: 22.36%    Mismatches:       98
Query Match:        6.31%       Indels:          108
BB:                  2          Gaps:              14

US-10-617-962-2 (1-1008) x H70219 (1-414)

QY      10 CAATTAAACCACTGATGATAGATGATATCCACCGCTGTAAGAAATGACGAGAGAT 69
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      178 GlnleerlelapegluaenglyAenlleleProVallelyshnFrhaenglyGln 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      70 ATAGTAGCTATACCA-----AACTTTAAGCAAAACAATGAGGTCATACAGCA 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 118 TCATATGGAATTGAATAT-----CGAGCTAAGAAATA 150  
| | | | | : : : : :  
| | | | | : : : : :  
Db 218 SerTyrLysValArgPheAspLeuIleProGluAenLysAsnIleAsnPhesgluLeu 237  
QY 151 ATATTAGCTTACGCTTGGCTGAAGTGATATTCAATATGATCTTAACCTTCGTATGAC 210  
| | | | | : : : : :  
| | | | | : : : : :  
Db 238 IleAsnValTyrTyrLeuAspGluLysAsnIleIleThr-----ProIleGlu 253  
QY 211 TATTATGAATAAAGAGACTGCTGAGAGATTATTCAAGATATATGCTATCTTTCA 270  
| | | | | : : : : :  
| | | | | : : : : :  
Db 254 TyrTyrLysAsnAsnIleAspMetSerProTyrTyrIleAsp----- 267  
QY 271 TCTGCACATATTAGTGAAAATGATGATCAAAATTTCAAGATATGCAAAATGCTTTTAT 330  
| | | | | : : : : :  
| | | | | : : : : :  
Db 268 -----LeuGlnGluAsnLysAspAsp-PheLeuLys----- 277  
QY 331 AAGATGAACCTGAGTATTGAAAGTCAATATCTCTCAAAACATTGGAAATGCTTCTGAGCTT 390  
| | | | | : : : : :  
| | | | | : : : : :  
Db 278 -----AlaIleLysIleLysLysGluTyrGlyLeuTyrIleGluLys 291  
QY 391 GAAAATTAACCATGAGTGCCTTATTCAGATGACGATTAATTTATGCACTATATTTTTC 450  
| | | | | : : : : :  
| | | | | : : : : :  
Db 291 sLys-LysGlnLeuGlnAsnLeuThrGluAsnAspLysLeu-----Asp 305  
QY 451 TCTGTACAGAAATTCACACTGAGAGAAAATCAACATCAATGCGCAGATTTTAA 510  
| | | | | : : : : :  
| | | | | : : : : :  
Db 306 AspPheLysGluPheLeuLeuLysAsnAsnAsn----- 316  
QY 511 TTAATGATTTCTTATTTACCTTATCTGCTGTAACCTCACTGGAAAGAGATTTTTC 570  
| | | | | : : : : :  
| | | | | : : : : :  
Db 317 -----IlePheSerLeuAsnThrIlePheSerAsnGlyAsnProIlePheThr 332  
QY 571 AAAAATTTTCAATGATGATTAGGCTAAATCATTAATGAGATTAATTAATGAGAAATA 630  
| | | | | : : : : :  
| | | | | : : : : :  
Db 333 -----TyrAlaIleAsnValLysAlaLysSerIleIleAsnTyrIleuIleThrLysGlu 350  
QY 631 CTTTCTMAACCTTCTTTCGACACCGCAGAGATTACCTGATGGCAGATAGATTATTTG 690  
| | | | | : : : : :  
| | | | | : : : : :  
Db 351 PheAsn----- 352  
QY 691 GCTGACCAACAGAAAGCGCTTAATAGAGAGTGAATTTAAAGACTTAATAACAA 750  
| | | | | : : : : :  
| | | | | : : : : :  
Db 353 -----IleAsnLeuThr 356  
QY 751 TCTAGGAATGATTTTCTAATATGAGAGGCGTCAAAACAAAGTATAGT---TCATTT 807  
| | | | | : : : : :  
| | | | | : : : : :  
Db 357 AsnGlnAsnSerGlnThrAlaLeuHisSerAlaIleIleGlnLysTyrAspLeuAsnPhe 376  
QY 808 ATAAAGAG---GTACAAAAGGTAAACGCTCCACAGACAGACGCAAAAGTATGGTACA 864  
| | | | | : : : : :  
| | | | | : : : : :  
Db 377 IleLysSerLeuIleGluLysGlyAlaAsnPro-----AsnIle 389  
QY 865 GCCAGTGCAGTACCTGGAATAATTCGCCGAATTAATTATATAGTGTGAGGCTAAGCCCA 924  
| | | | | : : : : :  
| | | | | : : : : :  
Db 390 ArgAspGlyAspAsn-----LysLeuProIleAsp---TyrSerAspLysThrSerGlu 406

RESULT 2  
A24594  
probable major surface antigen (83k, 19k, 42k) precursor - malaria parasite (Plasmodium  
C;Species: Plasmodium falciparum  
C;Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 09-Jul-2004  
C;Accession: A24594  
R;Holder: A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls,  
Nature 317, 270-273, 1985  
A;Title: Primary structure of the precursor to the three major surface antigens of Plasm  
A;Reference number: A24594; MUID:86014355; PMID:2995820  
A;Accession: A24594  
A;Molecule type: DNA  
A;Residues: 1-1640 <HOL>  
A;Cross-references: UNIPROT:P04933  
C;Superfamily: major merozoite surface antigen

C;Keywords: surface antigen

Alignment Scores:  
Pred. No.: 1.54 Length: 1640  
Score: 108.50 Matches: 86  
Percent Similarity: 34.96% Conservative: 50  
Best Local Similarity: 22.11% Mismatch: 136  
Query Match: 6.25% Indels: 117  
DB: 2 Gaps: 20

US-10-617-962-2 (1-1008) x A24594 (1-1640)

QY 121 TATGGAATTGAATATCGAGCT-----AAGAAATATATATAGCTTAC 162  
| | | | | : : : : :  
| | | | | : : : : :  
Db 205 PheAsnLeuLysIleArgAlaAsnGluLeuAspValLeuLysLysLeuValPheGlyTyr 224  
QY 163 GCTTTGCGTGAAGCGTAT---CATATGATATCAATCTTCGAGACTATATATAG 219  
| | | | | : : : : :  
| | | | | : : : : :  
Db 225 ArgLysProLeuAspAsnIleLysAspAsnValGlyLysMetGluAspTyrIleLysLys 244  
QY 220 AATAAGAGACTGCTGAG----- 237  
| | | | | : : : : :  
| | | | | : : : : :  
Db 245 AsnLysLysThrIleGluAsnIleAsnGluLeuIleGluLysSerLysThrIleAsp 264  
QY 238 -----AGAAATTTACAA---GAATATATG 258  
| | | | | : : : : :  
| | | | | : : : : :  
Db 265 LysAsnLysAsnAlaThrLysGluGluLysLysLysLysLysLysLysLysLysLysLys 284  
QY 259 TCTAATCTTTCATCTGCACCTATTAGTGAAAATGCTGATCAAAATTTCT---AAAGAT 312  
| | | | | : : : : :  
| | | | | : : : : :  
Db 285 LeuSerIleTyrAsnLysGlnLeuGluLysAlaHisAsnIleIleSerValLeuGluLys 304  
QY 313 ATGCCAAATGCTTTTATTAAGATGA-----CTGAGT--- 345  
| | | | | : : : : :  
| | | | | : : : : :  
Db 305 ArgIleAspThrLeuLysLysAsnGluAsnIleLysGluLeuLeuAspLysIleAsnGlu 324  
QY 346 TTGAGAGTCAATATCCCAAAACATTGGAAATGCTTCTGAGCTT-----GAAAT 396  
| | | | | : : : : :  
| | | | | : : : : :  
Db 325 IleLysAsnProProIleAsnSerGlyAsnThrProAsnThrLeuLeuAspLysAsn 344  
QY 397 AAACCATGAGTGCCTTATTCAGATGACGATTAATTAATAGCACTATATTT---TTCTCT 453  
| | | | | : : : : :  
| | | | | : : : : :  
Db 345 LysLysIleGluGluHisGluLysGluLysLysLysLysLysLysLysLysLysLysLys 364  
QY 454 GTACAGAAAT---CACTG----- 471  
| | | | | : : : : :  
| | | | | : : : : :  
Db 365 IleAspSerLeuPheThrAspProLeuLysGluLysLysLysLysLysLysLysLysLys 384  
QY 472 -----GAGAAAATCAACATCA 489  
| | | | | : : : : :  
| | | | | : : : : :  
Db 385 AsnIleAspIleSerAlaLysValGluThrLysGluSerThrGluProAsnGluTyrPro 404  
QY 490 AATGCCGAAGATT---TTTAATTAATGATTTCTTATTTACCTTACCTGCTGTAATCT 546  
| | | | | : : : : :  
| | | | | : : : : :  
Db 405 AsnGlyValThrTyrProLeuSerTyrAsnAspIleAsnAsnAlaLeuAsnGluLeuAsn 424  
QY 547 TCACGTGCAGAGAGATTTT---TCAAAAACCTTTTAA 582  
| | | | | : : : : :  
| | | | | : : : : :  
Db 425 SerPheGlyAspLeuIleAsnProPheAspTyrThrLysGluProSerLysAsnIleTyr 444  
QY 583 AATGATTAAGAGGCTAAATCAATTAGCAATTAATTTAGAGAAAA---AACTTTCTTAA 639  
| | | | | : : : : :  
| | | | | : : : : :  
Db 445 ThrAspAsnGluArgLysLysPheIleAsnGluLysGluLysLysLysLysLysLysLys 464  
QY 640 CCTTTCTTTCAGACACCGCAGAGATTACCTGATGCGAGATATAGTTATTTGCTGAGCA 699  
| | | | | : : : : :  
| | | | | : : : : :  
Db 465 LysLysIleGluSerAspLysLysSerTyrGluAspAspSerLysSerLeuAsnAspIle 484  
QY 700 ACAGAAAGCGCTAAATGAGAGAGTATTAAAGAACTTAATAAAATCAAGAAAT 759  
| | | | | : : : : :  
| | | | | : : : : :  
Db 485 ThrLys-----GluTyrGluLysLeuLeuAsnGluIleTyrAspSerLysPheAsnAsn 502  
QY 760 GGA-----TTTCTAATATGAAAGGCGCTGCACAAACAAAGTATGTTCATTTAA 813

```

Db      503 AsnIleAspLeuThrAsnPhcGluGlyMetMetGlybArgTyrSer-----Tyr 519
Qy      814 GAGGACAAAGGGGAGGACCTCCACAGCAGCAGGAAAGATTTGGACCCAGTGGC 873
Db      520 LysValGluLysLeuThrHisProAsnThrPheAlaSerTyr-----GluAsnSerLys 537
Qy      874 AGTAACTGGGAAATATGCGCAATATTTA-----TATAGTGTGAGGCTA 918
Db      538 HisAsnLeuGluLysLeuThrHisPheAlaLeuLysTyrMetGluAspTyrSerLeuTyr--- 556
Qy      919 AGCCAAAAGACAGGGTAACTTACTCAAAATGATGACAAATACATGACGGTTCAT 978
Db      557 -----AsnIleValValGlu 561
Qy      979 AGTGTGGAACCTCATTTATATAATATA 1005
Db      562 LysGluLeuLysTyrTyrLysAsnLeu 570

RESULT 3
AD1374
Internalin proteins, probable peptidoglycan bound protein (LpxTG motif) homolog lmo2396
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AD1374
R/Glaser, P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioeche
R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Kartz, U.
Science 294, 849-852, 2001
A/Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurupak, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluerer, T.; Simoes, N.; Tierce, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A./Title: Comparative genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AD1374
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-940 <GLA>
A/Cross-references: UNIPROT:O8V4N9; GB:NC_003210; PIDN:CADD00474.1; PID:G16411884; GSPDB:
A/Experimental source: strain EGD-e
C/Genetics:
A/Note: lmo2396

Alignment Scores:
Pred. No.: 1.63 Length: 940
Score: 108.00 Matches: 75
Percent Similarity: 37.03% Conservative: 52
Best Local Similarity: 21.87% Mismatches: 136
Query Match: 6.22% Indels: 80
DB: 2 Gaps: 16

US-10-617-962-2 (1-1008) x AD1374 (1-940)
Qy      112 ACAGCATCATATGAAATGAAATATGAGCT-----AGAAATATATATGACTTAC 162
Db      355 ThrValThrTyrLysAlaLysPheThrAlaThrSerSerLysGlyValProLeuSerTyr 374
Qy      163 GCTTGGCTGTAAGTGGTATTCATATATATATTAACCTCCGATGAC----- 210
Db      375 SerIleAsnValSerGlnProIleAsnValSerGlnGlnThrAspSerThrValSerVal 394
Qy      211 TATTATAGAAATTAAGAGAGCTGAGAGAAATTTTCAAGATATATATGCTAATCTTCA 270
Db      395 PheTyrGlnAspGluAsnGlyAsnGluLeuAlaProThrGlnThrLeuSerGlyLysSer 414
Qy      271 TCTGCATATATGAGTGAAGATGATGATCAAAATTTTAAAGATATGCAAAATGTTTAT 330
Db      415 -----GlyGluAspTyrGlnThrThrGluLysThrIleAlaAsn----- 427
Qy      331 AAGATGAAGTGAATTTGAAGTCAATATCTCAAAACATTGGAAGTTCCTGAGCT 390
Db      428 -----TyrGlnLeuLysGluIle 433
Qy      391 GAAATTAACCATGAGTGCCTTATTCAGATGACGATAAATATATATGACATATATTT--- 447

```

```

Db      434 GluGlyGlnAlaSerGlyGlnPheThrAspThrAspSerThrValThrTyrValTyrGlu 453
Qy      448 -----TTCTCTGTACAGAAATTCACATCGAGGAGAAATCAACATCA 489
Db      454 LysAlaAspGlyAlaProValThrValLysTyrValAspAlaAspGlyAsnAspLeuAla 473
Qy      490 AATCCCGCAAGATTTTAAATTAATGATTTCTTATTAATTAATTAATTAATTAATTAATTA 549
Db      474 ThrSerAspThrLeuAsnGlyLysIleAspAlaProTyrGlnThrSerAlaLysSerLeu 493
Qy      550 CTGGAGAGAGAGATT-----TTTCAAAAACCTTTTACATGAGATTA-----GAGCTAAA 600
Db      494 SerGlyTyrPheThrValLysThrThrProAsnAsnAlaThrGlyValPheThrAsnSerLys 513
Qy      601 TCATTAAGATTTATTTAGAGAGAAAAAATTTCTTAATTAATTAATTAATTAATTAATTAAT 660
Db      514 GlnThrValThrTyrValTyrGluLysAlaAspGlyAlaProVal-----ThrVal 530
Qy      661 AGATTACCTGATGAGAGATGATTATTTGCT-----GACCAACA 702
Db      531 LysTyrValAlaAspGlyAspGlyAsnGluLeuAlaThrSerAspThrLeuAsnGlyLysIle 550
Qy      703 GAACCGCTTAATGAGAGAGTGAATTTTAAAGAACTTAAAAAT----- 744
Db      551 AspAlaPro-----TyrGlnThrThrAlaLysSerLeuSerGlyTyrThrValLysThrThr 569
Qy      745 ---AACAAATCAGAAATGATTTTCTTAATTAATGAGAGCGCTGCAAAACAAAGTATGCT 801
Db      570 ProAsnAsnAlaThrGlyValPheThrAsnSerLysGlnThrValThrValTyrGlu 589
Qy      802 TCA-----TTTATAAAAGAGTACAA----- 822
Db      590 LysAlaAspGlyAlaProValThrValLysTyrValAlaAspGlyAsnGluLeuAla 609
Qy      823 -----AAGGTTAACGCTCCA-----CAGACAGCAGCAAAAGTAT 858
Db      610 ThrSerAspThrLeuAsnGlyLysIleAspAlaProTyrGlnThrThrAlaLysSerLeu 629
Qy      859 GGTACAGCAGTGGCAGTACCTGAGAAATTTCCGAAATATTTATATAGTGTGAGGCTA 918
Db      630 -----SerGlyTyrPheThrValLysThrThrProAsnAsnAlaThrGlyValPheThr 646
Qy      919 AGCCAAAAGACAGGGTAACTTACTCAAAATGATGACAAATGACAAATGACAAATGACAA 975
Db      647 AsnSerLysGlnThrValThrTyrValTyrGluLysAlaAspGlyAlaProValThrVal 666
Qy      976 CATAGTGT 984
Db      667 LysTyrVal 669

RESULT 4
T88317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C/Species: Melanoplus sanguinipes entomopoxvirus
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C/Accession: T88317
R/Alonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J.Virol. 73, 533-552, 1999
A/Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A/Reference number: Z20484; MUID:99102612; PMID:9847359
A/Accession: T88317
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1127 <AFO>
A/Cross-references: UNIPROT:O9VVT6; EMBL:AF063866; NID:G4049647; PIDN:AA097677.1; PID:G4
C/Genetics:
A/Note: MSV156

Alignment Scores:
Pred. No.: 1.65 Length: 1127
Score: 108.00 Matches: 72
Percent Similarity: 38.93% Conservative: 44

```



```

Qy 847 -----GCGAAGATATTGGTACACCGAGTGGCACTTAACCTGGAATAATTCGCGAAT 897
Db 1360 ProValProAlaAlaAlaAlaAlaAlaSerGly----- 1373
Qy 898 AATTATATAGTGGAGCTTAAGCCAAAAAGACAGGTTACTTACTCAAAATGATCT 957
Db 1374 -----SerAlaAlaThrThrlGluGluAlaAlaThrValAlaSerSer 1390
Qy 958 GACAAT 963
Db 1391 AspAsn 1392

RESULT 6
T41912
structural phosphoprotein - human herpesvirus 7 (strain J1)
C:Species: human herpesvirus 7
A:Variety: strain J1
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41912
R:Nicholas, J.
submitted to the EMBL Data Library, December 1995
A:Description: Determination and analysis of the complete nucleotide sequence of human h
A:Reference number: Z22022
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: T41912
A:Molecule type: DNA
A:Residues: 1-755 <N>
A:Cross-references: UNIPROT:P52519; EMBL:U43400; P1DN:AA054672.1
A:Experimental source: strain J1
C:Genetics:
A:Note: U11

Alignment Scores:
Pred. No.: 2.97 Length: 755
Score: 104.50 Matches: 71
Percent Similarity: 39.00% Conservative: 69
Best Local Similarity: 19.78% Mismatches: 106
Query Match: 6.02% Indels: 113
DB: 2 Gaps: 18

US-10-617-962-2 (1-1008) x T41912 (1-755)
Qy 13 TTAACACCTGATGATAGAGATATCCACCCGTTGAAGCAATAGACAGATATA 72
Db 109 LeuThrSerPheAsnValGlyGluThrGluPheValAsnValAlaSerPhe 128
Qy 73 GTACGTATATTAACCTTAAAGCAACAGATGAGGTCATACGATCATATGAAATTGAA 132
Db 129 LeuLysLeuGlySerPheLeuArgTrpGly-----ThrValThrHisAlaAlaSer 145
Qy 133 TAT-----CGAGCTAAG 144
Db 146 TyrValAsnLeuThrThrlGluGluArgAlaGluLeuGlyGluAsnLeuGlnLysAlaLys 165
Qy 145 AAAAATATATTAAGCTTACGCTTGGCTGTAAGTGTATATCATATGTAACCTTCT 204
Db 166 AsnAsnMetLeuSerPhe-----ThrIleYrGlnIleVal 177
Qy 205 GATGACTATATTAAGATTAAGACAGCTGAGAGAAATTTATCAA---GAATATATGCT 261
Db 178 AspProTpaAsnGluAsnGlyTyrTyrValThrAsnIleAsnArgLeuLeuThrGly 197
Qy 262 AATCTTTTCACTGCACCTATTAGTGAA-----AATGGTATCAATTTCTTAAATATG 315
Db 198 AsnLeuLeuIleThrLeuHisGlySerTrpMetAsnMetGluLysLeuAlaLeuAsnThr 217
Qy 316 GCAATGTGTTTATTAAGAATGACTGATTTTGAAGTCATATCTCAAAACATTTGG 375
Db 218 IleAsn----- 219
Qy 376 AATGTTCTGACCTGAAATAAACCATTTAGTCTTATTCAGATGACGATTAATATTA 435
Db 220 -----GluLysLysAsnAlaIleLeuLysAlaIleGluAsnAsnLysAsnPheVal 236

```

```

Qy 436 GCATTATATTTTCTCTGTACAGAAATTCACCTGAGGAAATCAACATCAATGCC 495
Db 237 SerIleYrSerYrGlnIleLeuSerLeuProLeuThrSerHisArgValThrSer--- 255
Qy 496 GCAAGATTTTAAATTAAT-----GATTCTTATTTACTTATCTGCTGTAACCTCA 549
Db 256 -----PhePheYrIleLeuThrGluAspPhe-----AspValIleThr--- 268
Qy 550 CTGGGAAGAGGATTTTTCAAAAACTTTTACATGATGATTAGAGGCTAAATCATAGAG 609
Db 269 -----LysSerLeuGln 272
Qy 610 AATTATATGAGAAAAAACTTTCTTAACCTTTCTTCCAGCCAGAGATTAACCT 669
Db 273 LeuHisAla-----LeuProValLysSerThrTrp 282
Qy 670 GATGCGAAGATAGTATTATTTGGCTGACCAACAGAGCGCTTAATGAGATGATTTT 729
Db 283 AspAspArgValLysPheThrProGluProIleGln-----ThrPhe 296
Qy 730 AAAGACTTAATAATTAACAAATCTAGAAATGATTTCTATATGAAAGGGGCTGCAAAA 789
Db 297 LysValLeu-----ThrAspLeuSerLysSerSerLeuSerAsnGlnPheGluSerSer 315
Qy 790 CAAAGTATAGT-----TCATTATATAAAGAGGTACAAAAG 825
Db 316 LysLysThrSerHisGlySerSerPheAsnProGluProPheIleLysThrGluGlnArg 335
Qy 826 GGTACCGCTCCACAGACAGACGAGCAAAAGTATGTGTAACGACAGTGGC-----AGTAAC 879
Db 336 SerAsnAsnThrLeuSerLysAspLeuPheValGlySerGluAspGlyLeuLeuSerSer 355
Qy 880 CTGCAAAA-----TTGCCGAATATTTATATAGTGTAGGCTAAGC 921
Db 356 ValLysLysAspSerMetIleLeuAspGluProArgAsnSerThrSerIleAsnAsnSer 375
Qy 922 CAAAA---GACAGGTAACCTTACTCA-----AATGATGCTACATACATCAATG 969
Db 376 LysLysMetHisArgIleLeuGlnThrGluIleLeuAspLeuThrAspGlnThrMet 394

RESULT 7
C85079
hypothetical protein AT4g08050 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85079
R:anonymous; The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: AB5001; MUID:20083488; PMID:10617138
A:Accession: C85079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1428 <STO>
A:Cross-references: GB:NC_001268; NID:g7267445; P1DN:CAB81142.1; GSPDB:GN00140
A:Map position: 4

Alignment Scores:
Pred. No.: 3.36 Length: 1428
Score: 104.00 Matches: 78
Percent Similarity: 35.04% Conservative: 59
Best Local Similarity: 19.95% Mismatches: 156
Query Match: 5.99% Indels: 98
DB: 2 Gaps: 16

US-10-617-962-2 (1-1008) x C85079 (1-1428)
Qy 16 ACACCTGATGATAGATGATATCCACCCGTTGAAGCAAAAT-----GCA 63
Db 750 ThrValGluGluArgArgAsnTyrProLeuGluGluGluArgValaTyrPheGluGluLys 769

```





649 CGACACCGCAGAGATTAAGTGGAGAG---ATAGGTTATTTGGTCGACCAAGAA 705  
||| ||| : : : : :  
Db 1940 AsnGlnProAen-----ProGluVallyAsnMetAenThrAspGluAsnAenThr 1957  
Qy 706 GCGCCTAA----- 714  
|||  
Db 1958 ThrThrLysLysLysAspAspAsnAspAsnGlnAsnAspPileTyrIleHisLeu 1977  
Qy 715 ---TGG-----AGAGTGAAGTTTAAAGAACTTAAATTAACAATCTAGG 756  
||| ||| : : : : :  
Db 1978 IleTyrAsnIleTyrAsnValArgThrLeuTyrThrGluArgLeuAsnAsnAspArgLys 1997  
Qy 757 AATGATTTTCTAATATGAGAGGCGCTCAAAACAAAGATATGATTCATTT----- 807  
Db 1998 -----IleAsnAen-----AlaLysArgLysTyrGluThrPheHisThrAsn 2011  
Qy 808 ----- 819  
||| ||| : : : : :  
Db 2012 MetAspAspIlePheAsnAspAspAsnAsnCysIleAsnIleIleAsnValGluAspAsn 2031  
Qy 820 CAAGAAGGTAAAGCTCCACAGACAGACAGCAAGAAATATTTGTCAGCCAGTGGCAGTAC 879  
||| ||| : : : : :  
Db 2032 LysGluGluAsnIleLysAspLeuLysTyrLysLysLeuLysThrAsnGluGlyGluLys 2051  
Qy 880 CTGGAAGAAA-----TTGCCGATATATTAATATATAGTGTGAGGCTAAGCCAA 924  
||| ||| : : : : :  
Db 2052 ValAspAsnGluPheIleGlnValThrAspAsnAsnIleIleGluIleAsnProLysLys 2071  
Qy 925 AAAGACAGGCTAACTTTACTCAAAATGATACTGAC---AATACATGACGGTTCATAGT 981  
||| ||| : : : : :  
Db 2072 Lys-----ThSerThrGlnAsnGluGluGlnProAsnIleAsnThrIleAsnGlu 2088  
Qy 982 GTTGAACCTCATTT 996  
||| ||| : : : : :  
Db 2089 AsnGlyAsnMetLys 2093

## RESULT 9

H97194  
uncharacterized ABC transporter, ATPase component CAC2392 [imported] - Clostridium acet  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: H97194  
R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: H97194  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-579 <KOR>  
A:Cross-references: UNIPROT:Q97GH5; GB:AE001437; PIDN:AAK80347.1; PID:G15025405; GSPDB:G  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2392

Alignment Scores:  
Pred. No.: 3.81 Length: 579  
Score: 103.00 Matches: 70  
Percent Similarity: 37.68% Conservative: 63  
Best Local Similarity: 19.83% Mismatches: 110  
Query Match: 5.94% Indels: 110  
DB: 2 Gaps: 16

US-10-617-962-2 (1-1008) x H97194 (1-579)

Qy 10 CAATTAACACCTGATGATGAAGTGATTCACACCGTGAAGAAAGCAATAGCAGAGAT 69  
||| ||| : : : : :  
Db 132 GlnAsnThrValGlyGluLeuAsnGlyLys-----IleGluGluIleIleGlyAsnGlu 209  
Qy 70 ---ATAGTACGTATATCAAACTTAAAGCAACAGATGAGGTCATATACAGCATATATGA 126  
||| ||| : : : : :  
||| ||| : : : : :

Db 210 LysIleVallyLysLeu-----PheGly 216  
Qy 127 ATGATATATGAGCTTAAGAAATAATATATTAAGCTTACGCTTTGGCTGATGATTCAT 186  
||| ||| : : : : :  
Db 217 TyrGluLysArgGlyGlyLys-----LysPheSerGluIleAsnSerArgLeuTyr 233  
Qy 187 AATGATCTAAACTTCTGATGATATTAAGATTAAGAAAGACTGCTGAGAAATTTAT 246  
||| ||| : : : : :  
Db 234 AsnGlyGluLeuAlaGlnPheTyrSerSerLeuThrAsnProSerThrArg----- 251  
Qy 247 CAAGATATATATGTTAATCTTTCATCTGACATATTAGTGAAATGCGATCAATTTCT 306  
||| ||| : : : : :  
Db 252 -----PheValAsnAsnIleThrTyrValLeuValGlyAlaValGlyGlyLeuLeuAla 269  
Qy 307 -----AAAGATATGCGCAATGTTTATTAAGATGAACTGATTTGAAGTCAATAT 360  
||| ||| : : : : :  
Db 270 ValLeuSerGlyLeuSerIleGlyThrIleSerSerPheLeuThrTyrSerThrGlnPhe 289  
Qy 361 CCTCAAAACATTTGGAATGTTCTGAGCTTGAATAATAACCATGAGCTTATTCAGAT 420  
||| ||| : : : : :  
Db 290 SerGlnProIleAsnAsnValThrGlyValAlaThrGlnLeuGlnAlaPheAlaSer 309  
Qy 421 GACGATTAATATTAATGACATATATTTTCTCTGACGAAATTCACCTGAGAGAAAT 480  
||| ||| : : : : :  
Db 310 AlaGluArgValPheSerIle-----LeuAspGluIlePro-----GluLys 323  
Qy 481 CAACATCAAAATGCGCGAAGATTTTAAATTAATGATTTCTATTTACTATTCATTCGCT 540  
||| ||| : : : : :  
Db 324 LysAspAsnGluGlyAlaLysLysPheGluHisCysGluGlyAsnIleSerPheAsnAsn 343  
Qy 541 GTAACTTCACCTGAGAGAGGATTTTTCACAAATCTTTTCAATGATTAAGAGCTTAA 600  
||| ||| : : : : :  
Db 344 ValSer-----PheSerTyrAsnLysLysGlnProLeu----- 354  
Qy 601 TCATTAAGAAATTAAT--ATTGAGAGAAAAAACTTTCTTAACCTTTCTTTTCGACACCG 657  
||| ||| : : : : :  
Db 355 ---IleGluAsnPheSerValAspIleLysLys----- 364  
Qy 658 CAGAGATTACTGANTGGAGAAATAGTTATTTGCGTGGACCAAGAGCGCTTAATG 717  
||| ||| : : : : :  
Db 365 -----GlySerThrIleAlaIleValGlyProThrGlyAlaGlyLysThr 379  
Qy 718 AGATG----- 723  
||| ||| : : : : :  
Db 380 ThrMetValAsnLeuLeuMetArgPheTyrAspIleAspGlyLysIleThrIleAsp 399  
Qy 724 -----AGTTTAAAGAACTTAAATAATCAATCTAGAAATGATTT----- 765  
||| ||| : : : : :  
Db 400 GlyLysAspIleAsnLysMetLysArgAsnAspValArgGlyGlnPheGlyMetValLeu 419  
Qy 765 ----- 765  
||| ||| : : : : :  
Db 420 GlnAspThrTrpLeuPheGluGlyThrIleLysGluAsnIleAlaTyrGlyLysProAsp 439  
Qy 766 -----TCTAATATGAGAGGCGCTGCAAAACAAAGTAT--AGTTCAATTTATAAA 813  
||| ||| : : : : :  
Db 440 AlaSerMetGluGluIleGluSerIleAlaLysLysAlaTyrIleHisAsnPheIleLys 459  
Qy 814 GAGGTACAAAGGCTTAACGCTCCACAGACAGACAGCAAGAAAGTATTTGTTACGCCAGTGGC 873  
||| ||| : : : : :  
Db 460 ArgLeuSerAspGlyTyrAspThrArgIle-----ThrGluSerGly 473  
Qy 874 AGTAACCTGGAAGAAATTCGCAATATTAATTAATAGTGG 912  
||| ||| : : : : :  
Db 474 GlyAsnLeuSerGluGlyGlnLysGlnLeuLeuThrIle 486

## RESULT 10

S60178  
gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy  
C:Species: Fusarium oxysporum  
C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: S60178  
R:Anaya, N.; Roncero, M.I.G.

Mol. Gen. Genet. 249, 637-647, 1995  
A>Title: SKIPpy, a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*.  
A.Reference number: S60178; MUID:96132549; PMID:8544829  
A.Accession: S60178  
A.Molecule type: DNA  
A.Residues: 1-853 <ANA>  
A.Cross-references: UNIPROT:Q00833; EMBL:UJ4658; NID:gs10695; PIDD:AAA8790.1; PID:gs106  
A.Experimental source: retrotransposon skippy  
C:Genetics:  
A.Mobile element: retrotransposon skippy  
C:Keywords: polyprotein

Alignment Scores:  
Pred. No.: 3.89 Length: 853  
Score: 103.00 Matches: 75  
Percent Similarity: 36.69% Conservative: 49  
Best Local Similarity: 22.19% Mismatches: 94  
Query Match: 5.94% Indels: 120  
DB: 2 Gaps: 18

US-10-617-962-2 (1-1008) x S60178 (1-853)

```
OY 193 TCTAACTCTGAT-----GACTATTATAGATATA 225
    |||||
DB 127 SerIysLeuProThrPheLeuThrGlnSerArgAlaPheIleThrTyTyrProAsnGln 146
    |||||
OY 226 -----GAGACTGCTGAG 237
    |||||
DB 147 PheArgAsnAspSerAlaIysValMetTyMetAlaGlyArgLeuIleLeysThrIalaIa 166
    |||||
OY 238 AGAATTATCA-----GATATATGCTCAAT-----ATG 315
    |||||
DB 167 GlnTrpPheGlnProIleMetAsnAspTyMetThrAsnProTyTyrLysLeuGlnPro 186
    |||||
OY 265 CTTTCACTGCACATATTAGTAAATGGGATCAAAATTTCTAAAGAT-----ATG 315
    |||||
DB 187 ArgThrAlaLeuLeuPheGlyIleAsnGlyArgHsIgluMetGluIleAlaLeuLysMet 206
    |||||
OY 316 GCAAAATGCTTTTATTAAGATGAATGATTTTGAAGTCAATATCTCAAAAACATTTGG 375
    |||||
DB 207 AlaPheGly-----ThrIleAspGlyLysGlyGlnAlaGluArgLysIleLys 222
    |||||
OY 376 AATGTTCCGAG-----CTTGAATAATACCATTTGAGTGT 411
    |||||
DB 223 ThrIleLysGlnThrGlySerAlaSerThrIleuGlyValGlnPheLeuGlnLeuAlaSer 242
    |||||
OY 412 TATTCAG-----GATGACGATTAATATATAGCATATTTTCTCGACAGGAAT 465
    |||||
DB 243 LysLeuProTrpAspGlnAspValLeuMetSerPhePhePheAspAlaLeuLysGlnGln 262
    |||||
OY 466 CCACGTGAGAAATCAACATCAATGCGCAGATTTTATTAATTGAATTTGATTTCTTA 525
    |||||
DB 263 ValGlnGlnIleuTrpGluLysAspArgProArg-----ThrIleuValGluTyr 279
    |||||
OY 526 TTTCACCTTCTGCTGTAATCTTGGAGAGAGATTTTTCAAAATACTTTTCAAT 585
    |||||
DB 280 -----IleAsnMetAlaValLysIleAspAspArgGlnPheAlaIleTrpArgHsAsn 297
    |||||
OY 586 GGATTAAGAGCTAAATCATTAAGAAATTAATTGAAGAAAAAATTTCTCAACCTTTC 645
    |||||
DB 298 SerArgGlyAsnLysGlyArgGlnAsp-----AsnLysPProArg 310
    |||||
OY 646 TTTCGACGACGACAG-----AGATTACCTGATGAGCAATAGATTAATTTGGCTGGA 696
    |||||
DB 311 TyHsValAsnGlnGlyArgThrArgGlnIleThrSerTyGlyThrGlnAlaGly 330
    |||||
OY 697 CCA-----ACAGAGCGCTAAATGAGAGTGAGTTT----- 729
    |||||
DB 331 PrometAlaIleGlyMetThrLysAspArgLysSerValIleThrCysTyIleAsnCysGly 350
    |||||
OY 730 -----AAGACCTTAATAATTAACAATCTAGAGATGATTTTCTAAT 771
    |||||
DB 351 LysLysGlyLysIleTyGlnArgGluCysLysAsnProValLysThrAsn----- 366
```

OY 772 ATGGAAGGGCGTCGAAAAAAGATAT-----AGTTCAATTATTAAGAAG 816
 |||||
DB 367 -----GlnLysTyTrpArgProValProGluGlyLysLysIleAsnMet 380
 |||||
OY 817 GTGCAAAAGGGTAACGCTCCACAGCAGCAGCAAGAAATATTTGTAACAGCGAGTGGCAGT 876
 |||||
DB 361 ValLysLysAsnGluGlnProGlnMetValIleLysTrpIle----- 394
 |||||
OY 877 AACCTGAAAAATTCGCCAATTAATTATAGTGAGCTTAAGCCAAAAAGACAGGTA 936
 |||||
DB 395 -----AsnMetThrArgLysAspLysTyr 402
 |||||
OY 937 ACCTTTACTCAA-----AATGATACCTGACAAATACAGACGCTTCAT 978
 |||||
DB 403 AspMetThrGlnAlaLysTyrlleAsnAsnPheAspProThrLeuAspValHis 420
 |||||

RESULT 11  
T09076  
hypothetical protein CG1 (strain HB3) - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
R:Su X.Z.; Kirtman, L.A.; Fujioke, H.; Welliams, T.E.  
Cell 91, 593-603, 1997  
A>Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistant  
A.Reference number: Z16556; MUID:98054002; PMID:9393853  
A.Accession: T09076  
A>Status: translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-950 <SUX>  
A.Cross-references: UNIPROT:O15788; EMBL:AF030690; NID:g2642510; PIDD:AAC47851.1; PID:g2642510  
A:Gene: cg1

Alignment Scores:  
Pred. No.: 3.92 Length: 950  
Score: 103.00 Matches: 82  
Percent Similarity: 35.14% Conservative: 48  
Best Local Similarity: 22.16% Mismatches: 122  
Query Match: 5.94% Indels: 118  
DB: 2 Gaps: 23

US-10-617-962-2 (1-1008) x T09076 (1-950)

```
OY 61 GCAGGAGATATAGTACGTATCACTTAACCAACAGATGAGGTGAT----- 111
    |||||
DB 540 SerGlyThrGlnThrLysValValAsnValPheGlnSerThrGlnCysHisIleTyrVal 559
    |||||
OY 112 -----ACAGCATCATATGAAATGGAATATGAGCTAAAGAAATTAATTAAGCTTACGCT 165
    |||||
DB 560 HisIleThrSerProTy-----ValCysAlaHisPro 570
    |||||
OY 166 TTG-----GCTTAAGTGGTATTCATTAAGTGA--- 192
    |||||
DB 571 LeuLysHisMetPrometLysSerLysAsnGlnThrValLysCysPheArgAsnIleTyr 590
    |||||
OY 193 -----TCTAACTCTGATGACATATTAAGATTAAGAGACTGAGAGAAAT 243
    |||||
DB 591 AlaAlaSerSerGlnLeuGlnAspLysTrpAlaSerAsn-----ThrPheAsnAsnIle 608
    |||||
OY 244 --TATCAAGATAT-----ATGTTAATCTTTTCATCTGACATTAATGAGTGAATGCT 294
    |||||
DB 609 MetTyArgAsnGlnTyAspSerMetProAsnIleIleSerSerAsnIleIleAsnAsnAsn 628
    |||||
OY 295 GATCAAAATTTCTTAAGATATGCAAAATGCTTTTATTAAGATGAACGTGATTTTGAAGT 354
    |||||
DB 629 AsnAsnVal-----AsnAsnPheTyArgAsn----- 636
    |||||
OY 355 CAATATCTCAAAACATTTGAGATGCTCTGAGCTTGAATAATTAACATGAGTGTCTAT 414
    |||||
DB 637 -----IleSerAsnLysIleTrpAsnProAsn-----IleSerSerTyr 649
    |||||
```

```

QY 415 TCAGAT---GACGATTAATTAATGACACTATTTTCTGTACAGGAAATTCACGTG 471
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 LysAsnHisAsnGluTyrThrLeuProAsnRheRheRheAsnGluAspLeuLeuLys 669
QY 472 GAGGAAATTCACAAATCAAT-----GCT 495
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 AsnAspAspRheLeuSerAsnAspAspLysMetAsnLeuGluTyrTyrLeuLeuGln 689
QY 496 GCAAGATTTTAAATTAATGATTTCTATTTACCTTATCGCTGTAATTCACATGGGA 555
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 690 AsnLysIleRheLysAsnLysAsnRheLeu---SerRheGluAlaIleProLysAspThr 708
QY 556 AGGAGATTTTTCACAAATTTTACATGATGATGAGGCTTAATTCATTAAGAGATTAAT 615
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 709 GluIleIleRheGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 723
QY 616 ATTGAGAGAAAAAATTTTCAAA-----CCTTTCTTTCGACACCGACAGATTAATCT 669
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 724 Val---ArgLysArgGluLysLysSerIleProIleRheHisIleGluAsnValAlaArg 742
QY 670 GATGGCAGATTAAGTTATTTGCT-----GGACCA 699
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 743 HisLysTyrTrpAsnTyrGlnAlaValIleSerTrpAspTyrSerCysRheAlaPro 762
QY 700 ACAGAAAGCGCTTAATG---AGAGTGAATTTTAAGAATCT----- 738
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 763 AsnGlu-----TrpLysLysSerRheSerGluTyrTrpProGluRheGlnAsn 779
QY 739 -----AAAAATCAAAATCTAGAAATGATTTTCT 768
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 780 ThrValHisTyrLeuLeuLeuIleAsnAsnLysLysLysLysLysLysLysLysLys 799
QY 769 AATATGAAAGGCGCTGCAAAACAAAGTATGATCTATTTAAAGAGTACAAAGGCT 828
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 800 TyrThrLysAspProAsnLys-----LysAsnCysAsnLysGlu 812
QY 829 AACGCTCCACAGACAGCAGCAAGATATGCTGACAGCCAGTACCTGGAATAA 888
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 813 AsnLysGluHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 831
QY 889 TTGCGCAATTAATTA-----TATAGTGAAGCTTAAGCCAAAGAC 930
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 ValAsnValAsnValAsnValAsnGluLysAspTyrLysAspRheAsnGlnHisAsp 851
QY 931 AGGTTACCTTTACTCAAAATGATCTGAC 960
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 LysAsnArgTyrAsnLysAsnLysMetAsp 861

RESULT 12
T28160
hypothetical protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28160; S23688
R:Kun, J.F.; Hibbs, A.R.; Saul, A.; McCol, D.J.; Coppel, R.L.; Anders, R.F.
Mol. Biochem. Parasitol. 85, 41-51, 1997
A:Title: A putative Plasmodium falciparum exported serine/threonine protein kinase.
A:Reference number: Z20482; MUID:97262159; PMID:9108547
A:Accession: T28160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2510 <KUN>
A:Cross-References: UNIPROT:Q94658; EMBL:U40232; NID:g1658332; PID:g1658333; PTDN:AA8540
A:Experimental source: strain FCQ27/PNG
R:Kun, J.; Hesselbach, J.; Schreier, M.; Scherf, A.; Gysin, J.; Mattei, D.; Pereira da
Res. Immunol. 142, 199-210, 1991
A:Title: Cloning and expression of genomic DNA sequences coding for putative erythrocyte
A:Reference number: S23688; MUID:91376328; PMID:1896607
A:Accession: S23688
A:Molecule type: DNA
A:Residues: 241,'S',243-244,673-959,'R',961-977,'S',1493-1494 <KUD>
A:Cross-References: EMBL:X53019
C:Genetics:

```

```

A:Gene: FESP
A:Introns: 645/2
A:Keywords: surface antigen

Alignment Scores:
Pred. No.: 4.51 Length: 2510
Score: 102.50 Matches: 83
Percent Similarity: 40.90% Conservative: 63
Best Local Similarity: 23.25% Mismatches: 106
Query Match: 5.91% Indels: 105
DB: 2 Gaps: 20

US-10-617-962-2 (1-1008) x T28160 (1-2510)
QY 82 CTAACTTATGAAACAGATGAGGCTGATACAGATCATATGATTAATGACACT 141
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2161 IleSerTyrLysArgCysAsnGlu-----IleGlnSerRheGlu----- 2173
QY 142 AAGAAATTAATTAATGATTTGCTTTGCTGTAATGATTAATGATTAATGATTAAT 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2174 ---LysValGluLeuThrIleLysAsnIleGluSerGluIleRheLysLeu----- 2189
QY 202 CCGATGACTATTAATTAAGATTAAGACCTGACAGAAATTTTCAAGAAATTAATGCT 261
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2190 -----ArgAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2204
QY 262 AATCTTATGCTGACATTAAGTGAATAATGCTATCA----- 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2205 AsnValGlnAsnAspMetMetAsnProAsnLysAspTyrTrpHisSerArgLeuThrAsn 2224
QY 301 -----ATTTCTAAGATTAAGCA---AATGGTTTATTAAG----- 333
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2225 LysTyrLeuAlaCysIleLeuLysAspRheLysCysHisAspAspRheTyrAsnArgMetLys 2244
QY 334 -----AATGAACGTGATTTGAAGCTCAATATCTCAAAACATTTG 375
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2245 LysIleLysLysSerRheAsnLysTyrGluLeuProLeuAsnTyrSerAspGluTyrTrp 2264
QY 376 AATGCTCTGAGCTTGAATAATTAACATTAAGCTCTTATTCAGATGACATTAATTA 435
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2265 AspLeu-----LeuThrAsn-----LeuLeuAsnTyrValProSerGluArgLeuLeu 2280
QY 436 GCACATATAT-----TTTTTCT-----GTACAGAAAT 465
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2281 AlaCysGluValLeuGluHisAspRheSerGluHisAsnGluLysIleHisAsnIle 2300
QY 466 CCACTGAGAGAAATCAACAAATCAATGCCCCAAGATTTTAAATTAATTAATGATTTCTTA 525
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2301 ---IleLysGluAsnAsnHisAlaArgHisValGluPheRheLysAspGluThrRheCys 2319
QY 526 TTATCTTATGCTGCTGTAATCTTCACTGGAAAGAGATTTTTCAAAACCTTTTACAT 585
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2320 AspThrLeu----- 2322
QY 586 GGATTAGAGGCTAATATCATTAAGATTAATTAAGAAATAA-----CTT 633
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2333 -----LeuLysAsnTyrLysGlnGluLysGluLysAsnTyrIle 2336
QY 634 TCTAACTTTCTTTCGACACCGACAGATTAATCTGATGACAGAAATGATTAATTTGCT 693
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2337 SerArgProPheRheGlnAspRheLysLysLysLysLysLysLysLysLysLysLys 2354
QY 694 GGACCAACAGAAAGCGCTTAATGAGAGCTGATTTTAAGAACTTAATAAATCAAACT 753
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2355 -AspGlnLysLysAsnIleGlnSerTyrLysLysLysLysLysLysLysLysLysLys 2373
QY 754 AGGATGATTTTCTTAATATGAAGGCGCTCAAAACAAAGTATGCTATTAATAAA 813
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2373 eProLeuGluLeu-----IleArgAsnLysIleIleLeuAsnSerLysIle 2388
QY 814 GAGGTCAAAAGGTGACCTCCACAGACAGCAGAAAGATTAATGATTAAGCAAGCTGC 873
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2388 eGlnLysLysAsnValIleIle---MetIleAsnValMetIleIleAspThrThrAsnAsn 2407

```

QY 874 AGTAACCTGGAAAAATTCGCCAATAATTATATAGT-----GTGAGGCTTA 918  
|||||  
Db 2408 SerAenIIeAsnAsnAsnAsnAsnAsnAsnAsnCysSerAsnAsnValLeuArgHis 2427  
QY 919 AGCCAAAAGACAGGCTAACCTTACTCAAAAT-----GATACCTGACAAAT 963  
|||||  
Db 2428 AsnLysLysSerAspIle---PheAsnLysAsnIIleValAsnGlnIIeAsnAsnLys 2446  
QY 964 ACAATGACGGTTCATAGTGTG-----GAACTCATTTATATAA 1000  
|||||  
Db 2447 ThrIle-LysTyrIleIleIleLeuMetCysGluIleIleIleLys 2460  
RESULT 13  
T31107  
telomerase reverse transcriptase - Oxytricha trifallax  
C.Species: Oxytricha trifallax  
C.Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004  
C.Accession: T31107  
R.Bryan, T.M.; Sperger, J.M.; Chapman, K.B.; Cech, T.R.  
Proc. Natl. Acad. Sci. U.S.A. 95, 8479-8484, 1998  
A.Title: Telomerase reverse transcriptase genes in Tetrahymena thermophila and Oxytricha  
A.Reference number: Z20985; MUID:98337940; PMID:9671703  
A.Accession: T31107  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-1132 <BRY>  
A.Cross-references: UNIPROT:O76332; EMBL:AF060230; NID:G342795; PID:G342796; PIDN:AA3  
C.Genetics:  
A.Gene: TERT  
Alignment Scores:  
Pred. No.: 4.71 Length: 1132  
Score: 102.00 Matches: 73  
Percent Similarity: 41.30% Conservative: 67  
Best Local Similarity: 21.53% Mismatches: 99  
Query Match: 5.88% Indels: 100  
DB: 2 Gaps: 18  
US-10-617-962-2 (1-1008) x T31107 (1-1132)  
QY 181 ATTCAATATGATCTTAACCTTCTGATGCTATTAAGAAATTA----- 225  
|||||  
Db 405 IIVeAlAsnGlnLysLysPheAspPheLysTyrLysSerLysAsnCysProLeuPro 424  
QY 226 GAGACTGCTGAGAGATTTATCAAGAATATATGCTTAATCTTCACTGACCATTAAGT 285  
|||||  
Db 425 GUAATTTPLysAsnLeuLysSerPheLeuGluAsp-----AlaIleValSerGly 442  
QY 286 GAAAAATGCTGATCAAAATTTCTAAAGATATG----- 315  
|||||  
Db 443 GILeuaArgGlyGlnValPheArgGlnLeuPheGluTyrGlnGlnAspGlnArgGlnIle 462  
QY 316 -----GCAATGCTTTTATAAAGAAATCAATGCAATTTTGA 351  
|||||  
Db 463 SerAenPheLeuThrGlnPheValAlaAsnValPheProLysAsnPheLeu-----Glu 480  
QY 352 GGTCAATATCTCAAAACATTTGGAATGTTCTGAGCTTGAATAATTAACCATTTGAGTCT 411  
|||||  
Db 481 GlyLys---AsnLysLysIlePheAsnLysLysMetLeuGlnPheValLysPheAsnArg 499  
QY 412 TATTGAGATGACGATTAATTAATTAATTAATTTTCTGTCGACAGAAATTCGA--- 468  
|||||  
Db 500 PheGluSerPheThrLysIleSerLeuLeuAsnLysPheArgValAsnGlnValSerTrp 519  
QY 469 -----CTGAGAGAAAAATCAACAA-----TCAAAATGCCGACAGA 501  
|||||  
Db 520 LeuSerPheLysCysLysAspGlnLysLysPhePheMetAsnGlnLysAsnIleVal 539  
QY 502 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 555  
|||||  
Db 540 PhePheLysValLeuLysTrpValPheGluAspLeuAlaIleThrLeuMetArgCysTyr 559

QY 556 -----AGAGAGATTTT-----TCAAAAACCTTT 579  
Db 560 PheTyrSerThrGluLysAlaLysGluTyrGlnArgIlePheTyrTyrArgLysAsnIle 579  
QY 560 TACAATGATTTAGAGGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 618  
|||||  
Db 560 TrpAsnMetIleMetArgLeuSerIleAspAspLeuLysGlnAsnLeuLysGlnVal 599  
QY 619 GAGAGAAAAAACTT-----TCTAAACCTTTT----- 645  
|||||  
Db 600 GILysLysGluMetArgIlePheCysGluSerGlnAsnPheAlaProGlyLysLeuArg 619  
QY 646 -----TTTGACCA-----CCGACAGATTAACCTGAT 672  
|||||  
Db 620 LeuIleProLysGlyAspThrPheArgProIleMetThrPheAsnArgLysIleProAsn 639  
QY 673 GGCAGAAATAGTTATTTTGGCTGACACCAACAGAGGCTTAATGAGAGTGAAGTTTAA 732  
Db 640 GlnValGlyLysPheGlnSerArgMetThrThrAsnLysLysLeuGlnThrAlaHisMet 659  
QY 733 GAACTTAATAATAACAAATCTAGAAAT-----GGATTTTCT-----AAT 771  
|||||  
Db 660 MetLeuLysAsnLeuLysSerLysMetPheLysHisSerPheGlyPheAlaValPheAsn 679  
QY 772 ATGGAAGGGCTGCAGAAACAAAGATATGTCATTTATAAAGAGTCAAAAGGCTAAC 831  
|||||  
Db 680 TyrAspAspIleMetLys---ArgTyrGlnAsnPheValGlnLysTyrLysGlnIleAsn 698  
QY 832 GCTCCAGACAGACAGCAGAAAGATTTGGTACAGCAGTGCAGATTAACCTGAGAAATTTG 891  
|||||  
Db 699 SerProLysLeuTyrPheValAlaMet-----AspIleGluLysCys 712  
QY 892 CGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 948  
|||||  
Db 713 TyrAspAsnValAspCysGluArgVal-----ValAsnPheLeuGln 726  
RESULT 14  
AD1525  
probable cell surface protein (LPXTG motif) [imported] - Listeria innocua (strain C1011  
C.Species: Listeria innocua  
C.Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C.Accession: AD1525  
R.Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Biocker  
J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karsel, U.  
Science 294, 849-852, 2001  
A.Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkaz, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A>Title: Comparative genomes of Listeria species.  
A.Reference number: AB1077; MUID:21537279; PMID:11679669  
A.Accession: AD1525  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-657 <GLA>  
A.Cross-references: UNIPROT:O920S2; GB:AL592022; PIDN:CAQ95972.1; PID:q16413192; GSPDB:G  
A.Experimental source: strain C1011262  
C.Genetics:  
A.Gene: Iid0740  
Alignment Scores:  
Pred. No.: 4.99 Length: 657  
Score: 101.50 Matches: 56  
Percent Similarity: 38.22% Conservative: 43  
Best Local Similarity: 21.62% Mismatches: 81  
Query Match: 5.85% Indels: 79  
DB: 2 Gaps: 12  
US-10-617-962-2 (1-1008) x AD1525 (1-657)  
QY 76 CGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 135  
|||||  
Db 49 ArgAsnLeuAsnLysLysArgGluAsp----- 57

```

QY 136 CGAGTAAGAAATATATAGTTCACGCTTGGCTGAAGTGTATTCATATGTA--- 192
Db 58 -----IleThrLeuAlaAspValGluThrIleThrSerLeuAspProValGly 73
QY 193 ---TCTAACTTCTGATGACTAT---TATAAGATTAAGAGACTGCTGAGCAATT 243
Db 74 AlaSerIleProAspAsnIleThrAspTyrLysAsn-----Leu 87
QY 244 TATCAAGAAATATATCTATCTTCATCTGACATTCATAGTGAAATGATGCAATT 303
Db 88 ThrAspLeuTyrIleThrGlnGlyThrLeuThrGluValProGluSerIleGlyLeu 107
QY 304 TCTAAAGATATGCAAAATGCTTTTATTAAGATGAATGATTTTGAAGTCAATATCT 363
Db 108 LysLysLeuThrPheLeuSerPheTyrAsnAsnLysLeuThr-----GluPhePro 124
QY 364 CAATAATGGAATGTTCTGAGCTTGAATTAACATTCAGTCTTATTCAGATGAC 423
Db 125 ThrValValTyrAspLeuProAlaLeuAsnSer----- 135
QY 424 GATAAATATTAAGCATATATTTTCTCTGACAGAAATTCACCTGAG---GAAAT 480
Db 136 -----LeuLeuLeuGlnArgAsnAsnIleSerGluIleProGluGluIleThrAsn 152
QY 481 CAACATCAAAATGCCCA-----AGATTTTAAATTAATTAATGAT 519
Db 153 MetSerSerHisLeuSerSerLeuAspValArgAsnAsnAsnLeuIleSerIleProAsp 172
QY 520 TTTTATTTTACCTTATCTGCTGTAACCTTCACCTGGAGAGAGATT----- 564
Db 173 LysIlePheThrThrLysTyrAlaSerArgSerGlyLeuIleIleAspThrGluGly 192
QY 565 -----TTTCAAAAACCTTTTACATGATGATTAAG 594
Db 193 AsnGlnIleThrSerAspValProValAspTyrLeuAspAsnTyrAsnAsnGly----- 210
QY 595 GCTAATCATTAAGAAATTAATTAAGAAATAACAATCTTAACCTTTCTTGACCA 654
Db 211 GlyAsnMetLeuGlnAsnTyrAsnTyrArgGlnLys----- 222
QY 655 CCGCAGAGATTACCTGATGGCAGAAATAGGTTATTTGGCTGACCAACAGAACGCTTAA 714
Db 223 -----GlnAspGlnLeuValTyrLysGlyAspProIleValValPro--- 236
QY 715 TCGAAGAGGAGTTTAAAGAACTTAAATTAACAATCTGAATGAGATTTCTAT 771
Db 237 TyrLysThrAspPheLysGlnLeuThrProAspLysSerLysGlyLeuAlaSer 255

RESULT 15
S05603
Major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (sf
NAlternate names: gp195 surface antigen
CSpecies: Plasmodium falciparum
CDate: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
CAccession: S05603; S04850
R:Myler, P.J.
Submitted to the EMBL Data Library, April 1989
AReference number: S05603
AAccession: S05603
AMolecule type: mRNA
A:Residues: 1-1639 <MVL>
A:Cross-references: UNIPROT:P04933; EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R:Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasm
A:Reference number: S04850; MUID:89345116; PMID:2668887
AAccession: S04850
AMolecule type: mRNA
A:Residues: 1504-1639 <MVL>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F.1-19/Domain: signal sequence #status predicted <Sig>

```

```

F.20-1639/Product: major merozoite surface antigen #status predicted <MVL>
Alignment Scores:
Pred. No.: 5.25 Length: 1639
Score: 101.50 Matches: 82
Percent Similarity: 34.70% Conservative: 53
Best Local Similarity: 21.08% Mismatches: 137
Query Match: 5.85% Indels: 117
DB: 2 Gaps: 19
US-10-617-962-2 (1-1008) x S05603 (1-1639)
QY 121 TATGAATGATGATTCGAGCT-----AGAAATATATATAGCTTAC 162
Db 205 PheAsnLeuLysIleArgAlaAsnGluLeuAspValLeuLysLysLeuValPheGlyTyr 224
QY 163 GCTTGGCTGTAAGTGATTT---CATATGATCTTAACTTCTGACTATATATAG 219
Db 225 ArgLysProLeuAspAsnIleLysAspAsnValGlyLysMetGluAspTyrIleLysLys 244
QY 220 AATAAGAGACTGCTGAG----- 237
Db 245 AsnLysLysThrIleGluAsnIleAsnGluLeuIleGluGluSerLysLysThrIleAsp 264
QY 238 -----AGATTTATCAAA---GAATATATG 258
Db 265 LysAsnLysAsnAlaThrLysGluGluGluLysLysLysLysLysLysLysLysLys 284
QY 259 TCTAATCTTTCATCTGACATTAAGTGTAAGAAATGCTATCAATTTCT---AAAGAT 312
Db 285 LeuSerIleTyrAsnLysGlnLeuGluGluLysIleAsnLeuIleSerValLeuGluLys 304
QY 313 ATGCAAAATGCTTTTAAAGATGAA-----CTGGAT----- 345
Db 305 ArgIleAspThrLeuLysLysAsnGluAsnIleLysGluLeuLeuAspLysIleAsnGlu 324
QY 346 TTTGAGGCTCAATATCTCAAAACATTTGGAATGTTCTGAGCTT-----GAAAT 396
Db 325 IleLysAsnProProProAlaAsnSerGlyAsnThrProAsnThrLeuLeuAspLysAsn 344
QY 397 AAACCATTAAGTCTTATTCAGATGACATTAATTAATTAAGCATATATTT---TTCTCT 453
Db 345 LysLysIleGluGluIleGluLysGluLysLysLysLysLysLysLysLysLysLys 364
QY 454 GTACAGCAAAAT-----CCACTG----- 471
Db 365 IleAspSerLeuPheThrAspProLeuGluLeuGluLysTyrLysLysGluLysAsnLys 384
QY 472 -----GAGAAATATCAACATCA 489
Db 385 AsnIleAspIleSerAlaLysValGluThrLysGluSerThrGluProAsnGluTyrPro 404
QY 490 AATGCCGCAAGATT---TTTAAATTAATGATTTCTTATTTACCTTATCTGCTGTAAC 546
Db 405 AsnGlyValThrTyrProLeuSerTyrAsnAspIleAsnAlaLeuAsnGluLeuAsn 424
QY 547 TCACGTGGAGAGAGATTTT-----TCAAAAACCTTTAC 582
Db 425 SerPheGlyAspLeuIleAsnProPheAspTyrThrLysGluProSerLysAsnIleTyr 444
QY 583 AATGATTAAGAGGCTTAATCATTAAGATTAATTAATTAAGAGAAA---AACTTTCTTAA 639
Db 445 ThrAspAsnGluLysLysPheLysPheLysGluLysIleLysLysLysLysLysLys 464
QY 640 CTTTCTTTCGACACCGCAGAGATTACCTGATGCGAGATTAAGTTATTTGGCTGACCA 699
Db 465 LysLysIleGluSerAspLysLysSerTyrGluAspArgSerLysSerLeuAsnAspIle 484
QY 700 ACAGAAGGCGCTTAATGAAGAGTGAATTTTAAAGATTAATAAATAACAAATCTAGAT 759
Db 485 ThrLys-----GluTyrGluLysLeuLeuAsnGluIleTyrAspSerLysPheAsnAsn 502
QY 760 GGA-----TTTCTAATATGAAGGCGCTGCAAAACAAAGTATATGTTCAATTATAAA 813

```

```

Db      503 AsnIleAspLeuThrAsnPhcGluLysMetMetGlyLysArgTyrSerTyrLysValGlu 522
QY      814 GAGGTACAAAAGGTTACGCTCCACAGACAGCAGCGAAAGTATTGTTACAGCCAGTGGC 873
Db      523 LysLeuThrHisHisAsnThrPheAlaSerTyrGluAsn-----SerLys 537
QY      874 AGTAACTCTGGAAAAATTGCCGAATTAATTAA-----TATAAGTGAAGCTA 918
Db      538 HisAsnLeuGluLysLeuThrLysAlaLeuLysTyrMetGluAspTyrSerLeuArg--- 556
QY      919 AGCCAAAAAGACAGGTTACCTTTACTCAAAATGATACGACATACAAATGACGGTTCAT 978
Db      557 -----AsnIleValValGlu 561
QY      979 AGTGTGGAACCTCATTTATAAATATA 1005
Db      562 LysGluLeuLysTyrTyrLysAsnLeu 570

```

Search completed: November 21, 2004, 08:49:11  
 Job time : 54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 06:30:11 / Search time 3133 Seconds

(without alignments)  
11723.986 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1 atggttataaataaacacc.....ctcatataaaatatatga 1008

Sequence: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Scoring table: 32822875 seqs, 18219865908 residues

Searched: Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62.8	6.2	1101	CNS0039G	AL063921 Drosophila
2	61.2	6.1	1225	CNS0161D	AL061721 Drosophila
3	60.6	6.0	1101	CNS008EV	AL069706 Drosophila
4	59.4	5.9	996	CNS000FV	AL071063 Drosophila
5	57	5.7	1101	CNS0021J	AL061936 Drosophila
6	56.4	5.6	928	CNS000KX	AL071865 Drosophila
7	55.8	5.5	1101	CNS000EVL	AL069706 Drosophila
8	55.6	5.5	844	BX139987	BX139987 Drosophila
9	55	5.5	1101	CNS0039G	AL063921 Drosophila
10	54.6	5.4	1084	CNS0071NH	AL063921 Drosophila
11	54.2	5.4	561	CA856457	CA856457 PESTOACO
12	53.8	5.3	1101	CNS016HF	AL06749 Drosophila
13	53.6	5.3	924	BX398967	BX398967 Drosophila
14	53.4	5.3	1092	CNS020K7	AL175686 Tetraodon
15	52.4	5.2	1327	CL644705	CL644705 CH213-79D
16	51.8	5.1	1074	CNS015ZR	AL106113 Drosophila
17	51.6	5.1	987	CNS014PQ	AL104456 Drosophila
18	51.6	5.1	1011	CF469747	CF469747 P4-CS Pia
19	51.4	5.1	600	CG91559	CG91559 EST0111 E
20	51	5.1	1350	CG744271	CG744271 P036-4-CO
21	50.8	5.0	853	CNS023KH	AL179594 Tetraodon
22	50.6	5.0	489	CF181864	CF181864 ISOIGF I
23	50.6	5.0	519	CF181855	CF181855 ISOIGF I
24	50.6	5.0	976	CL516956	CL516956 SAIL_98_C

25	50.6	5.0	1101	9	CNS0006J	AL062049 Drosophila
26	50.4	5.0	567	4	BM165529	BM165529 EST568052
27	50.4	5.0	807	7	CL489403	CL489403 SAIL_523
28	50.4	5.0	905	8	AZ550256	AZ550256 ENTEF58TR
29	50.4	5.0	1003	6	CD390192	CD390192 AGRNCOURT
30	50.2	5.0	1146	9	CNS021G2	AL176643 Tetraodon
31	50	5.0	548	2	BF330152	BF330152 RC0-BN031
32	50	5.0	548	2	BF330145	BF330145 RC0-BN031
33	50	5.0	770	8	BH549511	BH549511 BOGFS24TR
34	50	5.0	832	8	BH391984	BH391984 AG-ND-138
35	50	5.0	928	9	CNS00DKY	AL071865 Drosophila
36	50	5.0	1133	9	CG752293	CG752293 P046-4-H1
37	49.8	4.9	717	8	BH957562	BH957562 odg07E10
38	49.8	4.9	746	8	BH591510	BH591510 BOGFS24TR
39	49.8	4.9	849	8	AZ546009	AZ546009 ENTEF53TF
40	49.6	4.9	461	4	B1424095	B1424095
41	49.6	4.9	924	4	CNS020K6	AL208909 Tetraodon
42	49.6	4.9	1047	9	CNS01429	AL103611 Drosophila
43	49.4	4.9	1101	9	CNS003BB	AL064089 Drosophila
44	49.2	4.9	521	1	AU264780	AU264780
45	49.2	4.9	550	9	CC952969	CC952969 BOIHJ75TR

## ALIGNMENTS

RESULT 1  
CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921 GI:4941778  
VERSION AL063921.1  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila> and <http://www.genoscope.cns.fr>  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source location/Qualifiers

1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR08K10"  
/clone\_11b="RPCI-98"  
/note="End : TET3"

ORIGIN  
Query Match 6.2%; Score 62.8; DB 9; Length 1101;  
Best Local Similarity 20.5%; Pred. No. 0.00098;

Query Match	6.1%;	Score 61.2;	DB 9;	Length 1225;
Best Local Similarity	30.2%;	Pred. No. 0.0023;		
Matches 189;	Conservative 124;	Mismatches 306;	Indels 6;	Gaps 1

QY		200	TTCCTGAGTACTAATAGAATAAGACGTGCAGAGAAATTACAAGATTTATGT	259
Dd		1207	TKATTTTTWMAWAMAMMAMAMAAAAMWTWTTATATAAAATATATMAITATTAATW	1144
QY		260	CTAATCTTCACTGCACATTTAGTGAAAAVGGTGATCAATTTCTAAGATATGCCAA	319
Dd		1147	MAAAMAWTTTATATATWATTTWATATAAAAATAATATMMWAMWAMAMAMAWA	1088
QY		320	ATGGTTTATTAAGAATGAAGTGGATTTTGAAGTCAATATCCTCAAAACATTGGAATG	379
Dd		1087	WWATWWWTWKZAAATATAAAARTMAATATATAAAATVAMAWMWAMWMAAAATATATTTT	1028
QY		380	TYCCTGAGCTGAANAATAACCATTTGAGTGCTTATTCAGATGACGATTAATATATAGAC	439
Dd		1027	TTTTTTTAAAAAATAAAATAAATATATWATTAAMWMTAAAAATAAAATAAWTTA	968
QY		440	TATATTTTTTCTGTACAGAAAATTCATGTGAGAAAATCAACATCAATCGCGCAA	499
Dd		967	TTTAAWMTTTTTTBRAAAMWTWTTTTTTTTT-----WTTTAAAMAATWTTATWTRT	914
QY		500	GATTTTAAATTAATTAATGATTTCTTATTAATCCTATCGCTGAACCTCACGCGAAGA	559
Dd		913	TWTRRRRBAATWMTTTTTTTTTTTTTTTTATATTTTTTTTTTHHTYYMTTHMHM	854
QY		560	GGAATTTTTCAAAAACTTTTACAATGATTAGAGGCTAAATCATTTAGAGATATATATG	619
Dd		853	MMMMMMHMMMGHBGGKGJTTTTTKTTTKTKTKTTGMGAVTTTAAATTTTTNRD	794
QY		620	AGAGAAAAAACCTTTCAAACCTTTCTTTGACACCGCAGAGATTAACCTGATGCAGAA	679
Dd		793	TTYARAAAMDPTWNKTXTTHTTTXMKMKGMNMMNMWVGWTRTXTKTXKKDDGRF	734
QY		680	TAGTTTATTTGGCTGCACACACAGACGCTTAATGAGAGATGAGTTTAAAGAACTTA	739
Dd		733	MGGWTDIDAAAAAATAAAAHADAPAAAARAARAGAGGGGAGGRRGARGGABARDW	674
QY		740	AAAAATAACAATCTAGAAATGAGATTTTCTATATGAGAGGGGCTGCAAAACAAAGTATA	799
Dd		673	AKAANGAKETAWTRBDARRRGAGGTAAATTTTTTAAARMGKRRTTAAAMWAAMA	614
QY		800	GTTTCATTTATAAAGAGTACAAA	824
Dd		613	WTTATTTATRBMAAAAAATAAAAAA	589

RESULT 3	CNS00EVL/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
	CNS00EVL	1101 bp	DNA	linear	GSS	04-JUN-1999							
	Drosophila melanogaster genome survey sequence T7 end of Bac:												
	BACR25B23 of RPr1-38 library from Drosophila melanogaster (fruit fly), genomic survey sequence.												
	AL069706												
	AL069706.1	GI:494849											
	GSS.												
	Drosophila melanogaster (fruit fly)												
	Drosophila melanogaster												
	Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Prelygota;												
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;												
	Ephyridioidea; Drosophilidae; Drosophila.												
	1 (bases 1 to 1101)												
	Genoscope.												
	Direct Submission												
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :												
	BP 101 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr												
	- Web : www.genoscope.cns.fr)												
	Determination of this BAC-end sequence was carried out as part of a												
	collaboration with the Berkeley Drosophila Genome Project (BDGP).												
	The BDGP is constructing a physical map of the Drosophila												



## ORIGIN

[illegible]

1101 TTWWTATWTTTWTTTTWTATWTTAWATATATAWTAATTTWTTATATATWTAWWTATA 10422

[illegible]

279 ATTAGTGAATAATGCTGATCAAAATTTCTAAGATATGCGAAATGCTTTTAAAGATGA 338

339 ACTGATTTTGAAGTCATATATCCCAAA---CATTTGAATGTTCCTGACCTGAAA 394

395 ATAAACCAATGAGTGCCTTATTCAGATGACGATAAATTATTAGCACTATATTTTTTCTCTG 454

[illegible]

801 TWTWTATTTATWTATATAMWWTWATWWTAWAAWTAATAATWATATATATWTWATTWT 742

b  
741 TWTTTAAATTTTTAAATTATAWTAATTW-TTWAATWTITTTTTTTTTTTTAAATTGWA 683

682 AATTATTTTAAATWAAATWATTTTAATAAATTAATTAATTTTAAATTTTATWA 623

695 GACCAACAGAAAGCCCTAATGGAGTGACTTTAAAGAACTTAAATAACAATCTA 754

755 GGATGGATTTCATATAT 773

RESULT 4  
INSS00FVH

**ORIGIN**

Best Local Similarity: 33.4%; Pred.NO. 0.0058;  
Matches 160; Conservative 77; Mismatches 242; Indels 0; Gaps 0;

502 AAAAAAAAAA CCGAAAAAAAAATAAAAAAAAAAAAAAAAARATMWAAATAATAATA 561

[illegible]

170 CTGTAGTGGTATTCAATATCTAATACCTTCCTGATGACTATTATTAAGATAAAGACA 229

230 CTGCTGAGAAATTTATCAAGATATATGTCTAACTTTCATCTGCACTATTAGGTGAAA 289

290 ATGCTGATCAATTTCTAAAGATATGGCAATGTTTTATAGAATGA CTGATTTTG 349

[illegible]

802 TTKAMWATWAAATWAAAAAAWAAAHWATTTAAAAAAWWTAWWTAWWTW 861

410 CTTATTGAGATGACGATAAATTATTAGCACTATATTTTTTCTGTACAGGAATTCCAC 469

[illegible][illegible]

```

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_lib="RPCT-98"
/note="end : T7"

```

```
/orga11b11="Drosophila 11b11"
/mo1 type="genomic DNA"
```

/db xref="taxon:7227"

```
/clone="BACR27A24"
```

```
/clone_11b="RpCI-98"
```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

ORIGIN

```
Query Match      5.6%; Score 56.4; DB 9; Length 928;
Best Local Similarity 29.3%; Pred. No. 0.028;
Matches 113; Conservative 91; Mismatches 182; Indels 0; Gaps 0
```

OY	183	TCATGTAATCTCAAACCTCCGAGACATATAAAGATAAAAGA CTGGCAGAGAT	242
Dd	484	TTATATTAAATTAATATGTATTAATAATTTTTTATCAWAAAAAAAATVTTTTWWTTTT	543
OY	243	TTATCGAAGATATATGTCTAAATCTTTCATCTGCACATATAGTGGAATVGTGATCAAT	302
Dd	544	TWAAAAATATATTTWTTTTTTTTTWTTWAMTTTTATATWTATWTATTAAMWTAMMAT	603
OY	303	TTCTAAAGATATGCCAAATGGTTTTTATPAGAGTAAC TGATTTGAAAGTCAATATCC	362
Dd	604	TTTATATMTAATATWMAAAATATWTATTAATAATATWTWTTWTATMTATTATAWMAAAWA	663
OY	363	TCAAAACATTTGGAATGTCTCGACTGAAATTAACAATGAGTGTATTTCAGATGA	422
Dd	664	AAMAAAAAAMAATAAAAAAAAATTTAAAAATAAAAAAMAAAAATTTAMTTTTTTTTTTTTTWT	723
OY	423	CGATPAATTTATTAGCATATATTTTTCTGTGACAGAAATTCAC TGGAGAAATCA	482
Dd	724	ATWTATMAATTAATAAATATWMMMMALATWTDGKMNNNNAWMMMMMMWMAAAAMWWW	783
OY	483	ACAACTCAATGCCGCAAGATTTTTTAAATTAATGATTTCTTATTTAACCTTATCTGCTGT	542
Dd	784	WMWMAAAAAAAMAAAAAAMAAAAAADDDDDDDDDMWKAAXKKKKKKKKKKKKKB	843
OY	543	AACTTCACTGGGAAGAGATTTTTT	568
Dd	844	GKKKTXTXTXKGARWTTTTTTTTT	869

RESULT 7				
CNS00EVL				
LOCUS		1101 bp	DNA	linear GSS 04-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC. BACR25923 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL069706			
VERSION	AL069706.1 GI:4949849			
KEYWORDS	GSS.			
SOURCE	Drosophila melanogaster (fruit fly)			
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscormorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101) Genoscope.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr			
TITLE	Web : www.genoscope.cns.fr)			
JOURNAL	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosegawa and Aaron Mamoser in Pieter de Jong's laboratory at the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2, cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library			
COMMENT				

and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

**Source**

```
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
```

```
/db_xref="taxon:7227"
```

```
/clone="BACR29B23"  
/clone 1ih="BPCT-98"
```

```
/note="end : T7"
```

ה  
ה  
פ  
מ  
ה

## ORIGIN

Query Match	5.5%;	Score 55.8;	DB 9;	Length 1101;
Best Local Similarity	31.4%;	Pred. No. 0.039;		
Matches 201;	Conservative 109;	Mismatches 328;	Indels 3;	Gaps 1

QY	184	CATATGATGATTAACCTCCGATGACGATATATGAAGATTAAGAGACCTGACGAGATTT	243
Dd	446	MMGMMATTTCTCAHTTMMGMMGMMATWTMMABAAAATTTATWATWAAAAMMMW	505
QY	244	TATCAGATATATGCTTAATCTTTCATCTGCACTATTAGTGAAAATGCGATCAATT	303
Dd	506	WATTTTTMMWWTATTTWTTWMMWTATWMAAAAAAATAATTTAAAMWATATWAT	565
QY	304	TCTAAGATATAGCAAAATGCTTTTATATAGAATGAACTGGATTTGAGGCAATATCCT	363
Dd	566	TAAWAAATTTAAWAAWWTATATTAATWTAATAATWATTAATATAAAAAAATTTTTTWA	625
QY	364	CAAAACATTTTGAATGTTCTGAGCTTGAAATATAACCATTGAGTGCTTAATTCAGATGAC	423
Dd	626	TAAATTTTATAATTTAATTAATTAATTAATTAATTAATWATTTMMWTAATTAATATTTWA	685
QY	424	GATTAATTTATGACATATATTTTCTCTGTGACAGAAATTCCTGAGGAAATATCA	483
Dd	686	AATWMAAAAAAAMAAAAAATAWMAWATATWATWATTAATAAAMATAAAAAWMA	745
QY	484	CAATCAATGCGCAGATTTTAAATTAATTAATGATTTCTTAATTAACCTATCTGCTGTA	543
Dd	746	TWAMWATATATWMAATATATWTTWMAATWMAWMTATATMAATMAAATMAAAMAWA	805
QY	544	ACTTCACTGGAGAGAGATTTTTCAAAAAAATTTCAATGATTAAGAGCTTAATCA	603
Dd	806	---TAAATATWATWATWMAAAMAWATWMAWATATWMAATAWMAAAAAATWTAATA	862
QY	604	TTAGAGATTTATTTGAGAGAAAAAACTTCTTAACCTTTCTTGACACCGCAGAGA	663
Dd	863	TWATWATWMAAAAAAATAWMAATTTWTTTMMWMTATATPAAAMWATWMAAAAAAA	922
QY	664	TTACCTGAGCAGAAATGTTATTTGGCTGGAACAACAGAGCGCTTAATTGAGAGTG	723
Dd	923	AAAAAAATTAAMWMMWTATATTTWTTATTTAAATWTAATWMAATTWMAATWTATWTT	982
QY	724	AGTTTAAAGACTTAATAAATTAACAAATCTAGAAATGATTTTCTAATATGGAAGGCGT	783
Dd	983	WTATWATATATWMTATWMAWMAATATTTATTAAMWTAATTTTAAAAATATATATWAT	1044
QY	784	GCAAAACAAAGTATGTTCAATTTATATAAAGATGACAAA	824
Dd	1043	AMWWTAWATATWMAWMAATTTWTTATATATWTAATAATWMA	1093

RESULT 8	844 bp	DNA	linear	GSS 13-MAR-2001
LOCUS	BX139987			
DEFINITION	Danio rerio genomic clone DKEX-99E7,			genomic survey sequence.
ACCESSION	BX139987			
VERSION	BX139987.1	GI:27971314		
KEYWORDS	GSS;			
SOURCE	Danio rerio (zebrafish)			
ORGANISM	Danio rerio			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;			

REFERENCE	AUTHORS	TITLE	JOURNAL
-----------	---------	-------	---------

Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 844)  
Humphray, S.J., Huckle, E. and Durh

**COMMENT**

Submitted 13-MAR-2003 The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [humnery@sanger.ac.uk](mailto:humnery@sanger.ac.uk) Unpublished  
This sequence was generated from the T7 end of BAC 99E7. 99E7 is part of the Danoloky BAC library created by R. Plasterk and N.V. Keygene. Further details: [http://www.sanger.ac.uk/Projects/D\\_reio/](http://www.sanger.ac.uk/Projects/D_reio/).

## FEATURES

**Source**

ORIGIN

Query Match	5.5%;	Score 55.6;	DB 9;	Length 844;
Best Local Similarity	51.2%;	Pred. No. 0.042;		
Matches 130;	Conservative 0;	Mismatches 124;	Indels 0;	Gaps 0;

Oy	175	AGTGGTATTCAAAATGTAACCTTCCTGCATGACTAATTAAGAATAAAGACCTGT	234
Db	359	AATGTAATTAATTAATGATAATTAATTAATTAATGAATTAATTAATTAATTAATTAAT	418
Oy	235	GAGAGAAATTTATCAAGAAATATATGTCCTAACTTTTCATCTGCACATAATGAGTAAATGGT	294
Db	419	GATAATTAATTAATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	478
Oy	295	GATCAAAATTTCTAAAATGATATGCGCAATGGTTTTTAAGAAGTAACTCGATTTTGAAAGT	354
Db	479	AATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	538
Oy	355	CAATATCTCAAAACATTGGAAATGTTCCCTGCAGCTGAAAAATTAACCATTTGAGTGTAT	414
Db	539	AATATATGATTAATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	598
Oy	415	TCAATGACGATTA	428
Db	599	AATATGATGATTA	612

## RESULT 5

CNS0039G

LOCUS	1101 bp	DNA	linear	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #			

**ACCESSION**

VERSION AL063921.1 GI:4941778

## KEYWORDS

**SOURCE**

## ORGAN 1

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

1 (bases 1 to 1101)

AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Journal of Genomics
VOLUME	1(1998)
ISSUE	1
PAGES	1-1
DATE	1998

**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segreff@genoscope.cns.fr](mailto:segreff@genoscope.cns.fr))

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

**Source**

```

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"

```

**ORIGIN**

Query Match 5.5%; Score 55; DB 9; Length 1101;  
Best Local Similarity 18.8%; Pred. No. 0.06;  
Matches 132; Conservative 255; Mismatches 303; Indels 3; Gaps 1

[illegible]

melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of

QY	343	GATTTCGAGGTCGAATTCCTCCGCAAAACATTTGGAAATGTCCTGAGCTTGAAAAATTAACA	402
Db	855	AATTATTTAAAAATATATTTTATTTTAAATATATATATATTTATTTATTTTAAAAAAAAT	796
QY	403	TTGAGTCCTAATTCAGATGACGATTAATTTATTTGACATATATTTTTCTGTACAGAA	462
Db	795	TAAATATTTTWWATATATATATATATTAATAAATATNTAATTAATTTATTTTWTAMAWATA	736
QY	463	ATTCCACTGGAGAGAAATCAACATTCAAATGCCGAAGATTTTTTAATTTATGATTC	522
Db	735	AWTTTTTTTTTAATTAATAATATATATTTTAAATTTTAAATTTATTTATTTATTTATTTW	676
QY	523	TTATTTACCTATTCGTGCTGTAACCTGACGTGGAGAGAGATTTTTCGAAAAACCTTTAC	582
Db	675	AATTTTATTTATTTTATTTTATTTTATTTATATATATTAATAAATAAATAAATAATTTATTTW	616
QY	583	AATGATTTAGAGCTTAATCATTTAGAGAT	612
Db	615	ATAATATGKKGKRGKRWMAAAAAATTTTATTTAWY	586
RESULT 11			
CaBS6457		561 bp	mRNA
LOCUS			linear
DEFINITION	CaBS6457	EST 17-DEC-2002	
ACCESSION	PFEST0080e04.y1	Plasmodium falciparum 3D7 gametocyte cDNA library	
VERSION	CaBS6457		
KEYWORDS	CaBS6457.1	GI:27158994	
SOURCE	EST.		
ORGANISM	Plasmodium falciparum 3D7		
REFERENCE	Plasmodium falciparum 3D7		
AUTHORS	Eukaryote: Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
	1 (bases 1 to 561)		
	Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,		
	Matta, M., Gibbons, L., Martin, J., Wylie, T., Dante, M., Theising, B.,		
	Bowers, Y., Hillborn, M., Ritzer, E., Bennett, J., Jentes, E., Ronko, I.,		
	Tsangareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,		
	Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,		
	Wagerston, R., Wilson, R. and Sibley, D.		
	Washu Plasmodium EST Project		
	Unpublished (2001)		
TITLE	Contact: L. David Sibley		
JOURNAL	Washu Plasmodium EST Project		
COMMENT	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	Library was constructed by R. Haywood. DNA sequencing by:		
	Washington University Genome Sequencing Center For information on		
	obtaining a clone please contact: L. David Sibley		
	(sibley@orcim.wustl.edu), Washington University		
	Seq primer: -40UP from Glbco		
	High quality sequence stop: 412.		
FEATURES	location/Qualifiers		
source	1..561		
	/organism="Plasmodium falciparum 3D7"		
	/mol_type="mRNA"		
	/db_xref="taxon:36329"		
	/dev_stage="gametocyte (stage III-V)"		
	/lab_host="DH10B (Genesig, Invitrogen, Inc.)"		
	/clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA library"		
	/note="vector: pluscript SK plus; Site 1: EcoRI; Site 2: XhoI; The library was constructed by R Haywood. cDNAs were synthesized from gametocyte poly(A) + RNA by oligo d(T) priming, size-selected and directionally cloned into the EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR lambda vector (Stratagene). The primary library was mass excised as phagemid using the Exsist helper phage (Stratagene). Clones were mass excised using the Exsist helper phage (Stratagene), the phagemids were precipitated		

with pEG 8000 and extracted with phenol/chloroform.  
Phagemid DNA was electroporated into DH10B cells. Clone  
Availability: David Sibley, Washington University."

## ORIGIN

Query Match 5.4%; Score 54.2; DB 6; Length 561;

Best Local Similarity 45.0%; Pred.No.0.084; Mismatches 248; Indels 0; Gaps 0;

Matches 203; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

179 GTATTCATATGATCTAACTTCTGATGATCTATTATTAAGAAAGAGCTGTGAGA 238  
15 GTGCTCAATGTGATCTCTCAAAAAGATCTATTATTAATGAAGTACTTTGG 74  
239 GAATTTATCAAGATATATGCTTAATCTTTCATCTGACATTTAGGTGAAAATG 298  
75 GTTTATATCAAGATATTAATAAAGATATATACAGAAATCTGATGATCAACTATA 134  
299 AAATTTCAAGATATGCGAAATGTTTTTAATAAGAACTGATTTTGAAGTCAAT 358  
135 AAAAAGGAATCAATATCAATATTAATTAACAAATGAGAAATATCATTCAGTAATA 194  
359 ATCCCTCAAAAGATTTGGAATGTTCTGAGCTGAAATTAACCATGAGTGTATTCAG 418  
195 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 254  
419 ATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 478  
255 ATTAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 314  
479 ATCAACATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 538  
315 CTAAATCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 374  
539 CTGTAATCTCACTGGAAGAGATTTTTCAAAACCTTTTACATGATTAAGAGCTTA 598  
375 CTGATGTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 434  
599 AATCATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 629  
435 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 465

## RESULT 12

CNS016HF/c 1101 bp DNA linear GSS 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION BACN16H14 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL106749

VERSION AL106749.1

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequence

BP 191 91006 EVRY cedex - FRANCE (E-mail : sege@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -

http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billard at CERN (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector

pBelOBAC11.

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACN16H14"  
/clone\_1ib="DrosBAC"  
/plasmid="pBelOBAC11"  
/note="end : T7"

## ORIGIN

Query Match 5.3%; Score 53.8; DB 9; Length 1101;

Best Local Similarity 27.4%; Pred.No.0.11; Mismatches 259; Indels 2; Gaps 1;

Matches 150; Conservative 136; Mismatches 259; Indels 2; Gaps 1;

421 GACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480  
1101 KADDDARTRTAAARAAAGARKWMAADTAATVAKWMTWVDVAAADRVGSVTTT 1042  
481 CAACATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540  
1041 TTTTWTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 982  
541 GTAACCTCACTGGAAGAGATTTTTCAAAACCTTTTCAATGATTAAGAGCTTAAA 600  
981 NTWTATAGAADARRRRRDMMKATTAAGGRRBAWRATWMAAARAKWMAAAMWMD 922  
601 TCATTAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 660  
921 TTTTAAARRRRGAAGWATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 662  
661 AGATTACCTGATGAGAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 720  
861 VSSDPAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 804  
721 GTGAGTTTAAAGAACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780  
803 KDTTATTTGCGGGRMTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 744  
781 GCTGCAAAACAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840  
743 WAMRAAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAAR 684  
841 ACAGAGCGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900  
683 ASSSTAAADGATGATGGAAGKGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 624  
901 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960  
623 TATTAMRRRTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 564  
961 AATACAA 967  
563 ATTAATA 557

## RESULT 13

BX398967

LOCUS BX398967

DEFINITION BX398967 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

clone CS0D1067YG04 3-PRIME, mRNA sequence.

ACCESSION BX398967

VERSION BX398967.2

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 924)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA library and normalization

Unpublished (2001)

JOURNAL On May 13, 2003 this sequence version replaced gi:30613827.

Contact: Genoscope

Genoscope - Centre National de Sequence

BP 191 91006 EVRY cedex - France



REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1327)	Kingsley, D., Grimwood, J., Dickson, M., Schmutz, J. and Myers, R.M	Expressed sequence tags from <i>Gasterosteus aculeatus</i>	Unpublished (2004)	Contact: Grimwood, Jane

FEATURES	Location/Qualifiers
source	1. .1327

## ORIGIN

Query Match	5.2%;	Score 52.4;	DB 9;	Length 1327;
Best Local Similarity	46.0%;	Pred. No. 0.24;		
Matches 230;	Conservative	0;	Mismatches 266;	Indels 4;
				Gaps 2

Qy	149	TAAATTAAGCTTAAGCTTGGTGGTGAAGGGAATCAATATGATCTAAACCTCCGATG	208
Db	776	TATTTTAAATTAATTTTAAAAATTAATTAATTAATTTAAAAATTTAATTTTATA	835
Qy	209	ACTATTAATGAATTAAGAGACTGCTGAGAGAAATTATCAAGATATATGTCTAACTTT	268
Db	836	AATTTATATTAAATTTTAAATTTTATATTTTAATTTTAAATTTTAAATTTTATATA	895
Qy	269	CATCGCACTATTAGGTAAAAATGTGATCAATTTCTTAAGATATGGCAATGGTTTTT	328
Db	896	TTTTTTTAAAAATTAAATTAATA--TATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	952
Qy	329	ATAAGATGAAGCTGGATTTTGAAGTCAATATCCCAAAACATTGGAAATGTTCCGAG	388
Db	953	TTAAAAATTAATTAATTAATTTATTAATTTAAATTTTAAAAA-AATTTTAAATTTTAAATA	1011
Qy	389	TTGAAAAATTAACCATGAGTGGCTTAATCAGATGACGATTAATTTATAGACATATATTTT	448
Db	1012	TTTTTATTTTAATAATTTTAATTTATTAATTAATTTTAATTAATTAATTTATTTTAAAA	1071
Qy	449	TCTCTGTACAGAAATTCACACTGAGAGAAATCAACAAATGCAAAATGCCGACAGATTTTAA	508
Db	1072	ATTTTATTTTATTTATTTAAAAATTAATTAATTAATTAATTTTATTAATTAATTAATA	1133
Qy	509	AATTAATTAATTTCTTAATTAATCTTATCTGCTGTAATTTCACTGGAGAGAGAAATTTT	568
Db	1132	AAATTAATTTATTTATTAATTAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTT	1199
Qy	569	CAAAAACTTTACATGGATTTAGAGCTAAATCATAGAGATTAATTTGAGAGAAAA	628

D <sub>b</sub>	1192	TNTTAATTAATTAAATAAATAATTTTTTTTAAAAAAAATATATTATATAAATTTATTTATNATATAT	1251
Q <sub>y</sub>	629	AACCTTCTAAACCCTTCTT	648
D <sub>b</sub>	1252	ATATATAAATAATTTTATTTT	1271

Search completed: November 21, 2004, 08:32:25  
Job time : 3138 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_n3p model

Run on: November 21, 2004, 08:32:30 | Search time 139.5 Seconds  
(without alignments)  
8315.092 Million cell updates/sec

Title: US-10-617-962-2  
1335  
Perfect score: 1 atggttatacaattacacc.....ctcatataaataatata 1008  
Sequence:

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1825181 segs, 575374646 residues

Total number of hits satisfying chosen parameters: 3650362

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Command line parameters:

-MODEL=frame+np\_model -DEV=xlp  
-O=/cgen2\_1/USPTO\_spool\_p/US10617962/runat\_17112004\_163922\_9749/app\_query.fasta\_1.1159  
-DB=Uniprot\_02 -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=200000000  
-USER=US10617962@cgen\_1\_1\_152@runat\_17112004\_163922\_9749 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

Uniprot\_02: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1716	98.9	335	2	Q7N4K5
2	124.5	7.2	1489	2	Q7RJ32
3	119.5	6.9	1485	2	Q7RCB8
4	116.5	6.7	487	2	O8ESK1
5	115.5	6.7	473	1	TIG PROMP
6	113.5	6.5	434	2	O8I226
7	113	6.5	682	2	O6F1M0
8	113	6.5	908	2	O8I5F7
9	112.5	6.5	1822	2	O9U0N4
10	111	6.4	995	2	Q7RT20
11	111	6.4	1794	2	O8IHV6
12	110	6.3	3468	2	O8I104
13	109.5	6.3	414	1	YQ28 BORBU
14	109.5	6.3	1491	2	Q7RN55
15	109	6.3	1119	2	Q7RLZ8
16	108.5	6.3	730	2	O86A28

17	108.5	6.3	795	2	Q71X29	Q71X29 listeria mo
18	108.5	6.3	795	2	AAT05136	AAT05136 listeria
19	108.5	6.3	900	2	O8IKV2	O8IKV2 plasmodium
20	108.5	6.3	1951	2	O8ILV5	O8ILV5 plasmodium
21	108.5	6.3	2393	2	Q7RCR2	Q7RCR2 plasmodium
22	108.5	6.3	3482	2	O8ID46	O8ID46 plasmodium
23	108	6.2	447	2	O6LX59	O6LX59 methanococ
24	108	6.2	447	2	CAF31049	CAF31049 methanoc
25	108	6.2	940	2	O8Y4N9	O8Y4N9 listeria mo
26	108	6.2	1127	2	O9YV6	O9YV6 melanoplus
27	108	6.2	2612	2	O8ISX5	O8ISX5 plasmodium
28	107.5	6.2	1092	2	O6FU29	O6FU29 candida glia
29	107.5	6.2	1171	2	O8I3F4	O8I3F4 plasmodium
30	107.5	6.2	2558	2	O8IKP6	O8IKP6 plasmodium
31	106.5	6.1	499	2	O6U7S2	O6U7S2 crinipellis
32	106.5	6.1	499	2	AAQ74285	AAQ74285 crinipell
33	106.5	6.1	867	2	O7REJ4	O7REJ4 plasmodium
34	106	6.1	353	2	O8I7S1	O8I7S1 plasmodium
35	106	6.1	412	2	Q7RT80	Q7RT80 plasmodium
36	106	6.1	1330	2	O8IBJ3	O8IBJ3 plasmodium
37	106	6.1	1636	2	O8ID77	O8ID77 plasmodium
38	106	6.1	1785	2	O25685	O25685 plasmodium
39	106	6.1	2740	2	Q7RES2	Q7RES2 plasmodium
40	105.5	6.1	460	2	O89627	O89627 clostridium
41	105.5	6.1	512	2	Q7RKX6	Q7RKX6 plasmodium
42	105.5	6.1	744	2	Q7RS98	Q7RS98 plasmodium
43	105.5	6.1	750	2	O6MSZ0	O6MSZ0 mycoplasma
44	105.5	6.1	750	2	CAE77248	CAE77248 mycoplasma
45	105.5	6.1	755	2	O56271	O56271 human herpe

#### ALIGNMENTS

##### RESULT 1

ID Q7N4K5 PREDIMINARY; PRT; 335 AA.  
AC Q7N4K5;  
DT 01-MAR-2004 (TREMblrel. 26, Created)  
DT 01-MAR-2004 (TREMblrel. 26, Last sequence update)  
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
DE Unknown protein.  
GN OrderedLocusNames=pu12326;  
OS Photorhabdus luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=141679;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T01;  
RX MEDLINE=22957627; PubMed=14528314;  
RA Duchaud E., Rusniok C., Frangoul L., Buchrieser C., Givaudan A.,  
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,  
RA Daes B., Derose R., Derzelle S., Freysinet G., Gaudreau S.,  
RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,  
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;  
RT "Type genome sequence of the entomopathogenic bacterium Photorhabdus  
RT luminescens.";  
RL Nat. Biotechnol. 21:1307-1313(2003).  
DR EMBL; BX571866; CAE14619.1; -.  
KW Complete proteome.  
SQ SEQUENCE 335 AA; 38038 MW; ASE7P93A73E150B0 CRC64;

##### Alignment Scores:

Pred. No.: 1,26e-125 Length: 335  
Score: 1716.00 Matches: 331  
Percent Similarity: 99.10% Conservative: 1  
Best Local Similarity: 98.81% Mismatches: 3  
Query Match: 98.90% Indels: 0  
DB: Gaps: 0

US-10-617-962-2 (1-1008) x Q7N4K5 (1-335)

QY	1	ATTCGTTATACATTTAAACACCTGATGATAGAAAGGAGATATTCACCCGTTGAAAAAGCAATA	60
Db	1	MetValIIEgInLeuThrProaspPaspSergILTyIreProValGIuIySgInIle	20
QY	61	GCAGAGATATATGTACGTATACCTTAACCTTTAAGCAAAACAGATAGAGGGTCTATACGATCA	120
Db	21	AlaGIyAspIleValAlrGIIEuEanPheIySgInThrAspGIuIyAlsrnAlaSer	40
QY	121	TATGGAAATTTGAATATTCGAGCTTAAGAAAAATATATTAAGCTTACCGCTTGGCTGAAGTGT	180
Db	41	TyrGIyIIEgILTyAlrGIAlaIySgIIEIIEuAlaIyAlrAlaIeulAlaValSergIy	60
QY	181	ATTCAATATGTATCTTAACCTTCTCTGATGACTATATTATAGAAATTAAGAGACTGCTGAGAGA	240
Db	61	IIEhIsanValSerIySLeuProaspPyrTyIlySasnIySgIuThrAlaGIuIyArg	80
QY	241	ATTATCAAGATATATGCTPAATCCTTATCTGACCTATTAGTGAAGAAATGGTGAATCAA	300
Db	81	IIEIyrgInGIuIyIwCeserSasnIeuserSerAlaIeueuGIyGIuSngIyAspGIu	100
QY	301	ATTTCTAAAGATATAGGCAAAATGGTTTTTAATAGAAATGAATCTGGATTTTGAAGGCTCAATAT	360
Db	101	IIEserIySAspMeAlaSnGIyPheTyIryIyAsnGIuIeunAspPheGIuGIyGIuTyI	120
QY	361	CCTCAAAACATTTGGATATGTTCTGAGCTTGAAATTAACATTAGTGTCTTATTCAGAT	420
Db	121	ProGIuAsnIIEIryAsnValProGIuIeunGIuSasnIySProIeuserAlaTyIySerAsp	140
QY	421	GACGATTAATTTATTTGGCACTATATTTTCTCTGTACAGAAATTTCACTCGAGGAGAAAT	480
Db	141	AspAspIySLeuEunAlaIeutyIryPhePheSerValGIuGIuIIEProIeunGIuGIuAsn	160
QY	481	CAACATCAAAATGCGCAAGATTTTAAATTAATATGATTTCTTATTTACCTTATCTGTCT	540
Db	161	GIuGIuIeSerAsnAlaIaIarPhePheIySLeuIIEAspPheIeunIeuthrIeuserAla	180
QY	541	GTAATCTTCACTGGGAAAGAGATTTTTCAAAAACCTTTATACATGATTAAGAGGCTAAA	600
Db	181	ValThrSerIeugILyAlrGIAlaIySgIIEPheSerIySAsnPheTyIySngIyIeugIuSerIyS	200
QY	601	TCATTATAGAAATTAATTTAGAGAAAAAACTTCTAAACCTTCTTGGACGACCGGACG	660
Db	201	SerIeugIuAsnTyIryIIEGIuAlrGIyIySAspPheProIySProPhePheArgProProGIu	220
QY	661	AGATTACCTGATGGGCAAGATAGGTTATTTGGCTGAGCAACAAGAGCGCTTAATGGAGA	720
Db	221	AlrGIeunProAspGIyAlrGIIEGIyIryIeunAlaGIyProthrGIuAlaIyProIySAsp	240
QY	721	GTGAGTTTAAAGAACTTAATAAATAACAATCTTAGAATGAGATTTTCTAATATGAAAGG	780
Db	241	ValSerPheIySgIuIeunIySAsnAsnIySerAlrGIAsnGIyPheSerAsnMetGIuGIy	260
QY	781	GCTGAGAAACAAAGATATGTTCAATTTATTAAGAGGTACAAAAGGTAAACGCTCCACAG	840
Db	261	AlaAlaIySgInIySerIySerPheIIEIySgInValGIuIySgIyAsnAlaIyProGIu	280
QY	841	ACAGCAGCGAAAAGTATTTGTCACAGCGAGTGGAGTAACTCTGAAAAAATGGCGGAATAT	900
Db	281	ThrAlaAlaIySgIIEGIyThrAlaSerGIySerAsnIeunGIuIySLeuProAsnAsn	300
QY	901	TTATATAGTGTAGGCTTAAGCCAAAAAGAACGGGTAACTTTAATCTCAAAATGATACTGAC	960
Db	301	LeuTyIySerValAlrGIeunSerGIuIySAspArgValThrPheThrGIuAsnAsnAlaPThrAsp	320
QY	961	AATACAAATGACGGTTCATATAGTGTGGAATCTCATTAATAAATATA	1005
Db	321	AsnThrMetThrValIIEsSerValGIyThrAlaTyIryIySasnIle	335
RESULT 2			
ID	Q7R3J2	PRELIMINARY	PRT; 1489 AA.
AC	Q7R3J2		
DT	01-MAR-2004 (TREMBLrel, 26, Created)		

[illegible]

QY 517 GATTTCCTATTT---ACCTATCTGCTGTAACTTCACTGGAAGAGGATTTTTCANAA 573  
 Db 524 GLeuileIeGnIProthIleGlnAsnleuIeSerThngIngluGluLeu---Lys 542  
 QY 574 AACTTTTCAATGATTTAGAGCTTAATCAATTAGAAATTATTATTGAGAGAAAAACTT 633  
 Db 543 AsnIleGlnHsIglAsnLysAlaLysSerTyLeuAspTyValValGluAsnGluPhe 562  
 QY 634 TCTAAACCTTTCTTTCGACACCCGAGAGATTACTAGGAGAAAT- 681  
 Db 563 AspHglIeValIthnIleuLysLysLysLysLysLysLysLysLysLysLysLys 582  
 QY 682 GGTATTGCTGTCGACCAACAGAGCGCTTAATGAGAGTGTATTTAAAGAACTTAA 741  
 Db 583 GluTyLeuLysAlaAsnGluGlyPheAspAsnIleSerLysSerIleAsnAsnValLys 602  
 QY 742 AATAAATAATGAGATGATTTTCTAATATGAGAGGGCTGCAAAACAAATATAGT 801  
 Db 603 AsnSerThrAspGluAsnSerLeuAsnIleLeuAsnGlnThrLysGlnMetTyCtu 622  
 QY 802 TCATTATTAAGAAGGTACAAAAGGT-----AAC 831  
 Db 623 AsnIleValSerLysThrTyAsnSerTyLysAspAlaGluAsnIlePheIleAsn 642  
 QY 832 GCTCCACAGACAGACGCGAAA-----AGTATTGCTACAGCCAGTGCAGTACCTG 882  
 Db 643 IleProLysLysAlaAsnSerLeuAsnIleGlnIleLysSerSerSerCylIleAsnLeu 662  
 QY 883 GAAAAATTCGCGAAT-----AATTATATAGTGTAGAGCTTAAGCCAAAAGACGGTA 936  
 Db 663 PheLysAsnIleAsnIleAlaIleLeuProTyLeuAspSerGlnLysAspThrLeu 682  
 QY 937 ACCTTACTCAAAATGATATGACAAATGACAGGCTTCAATGCTTGAAGTCAATAT 996  
 Db 683 ThrPheIleProSerProGlnLysThrSerGluThrTyThrLysIleSerAspSerTy 702  
 QY 997 AAAAATATA 1005  
 Db 703 AsnThrLeu 705  
 RESULT 3  
 QYRCB8 PRELIMINARY; PRT; 1485 AA.  
 AC QYRCB8;  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE RNA recognition motif, putative.  
 GN Name=PY05866;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNL;  
 RX PubMed=1236865;  
 RA Carlson J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Petrea M.,  
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shalom S.J., van Aken S.B., Riedmuller S.B., Feldblum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,  
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,  
 RA van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii.";  
 RU Nature 419:512-519 (2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBD whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL, AABL01001921; EAA17961.1; -.

DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; RRM 1; 1.  
 DR PROSITE; PS0102; RRM; 1.  
 DR PROSITE; PS00030; RRM RNP 1; UNKNOWN 1.  
 SQ SEQUENCE 1485 AA; 172775 MW; E12120E6F2E1CB CRC64;  
 Alignment Scores:  
 Pred. No.: 0.852 Length: 1485  
 Score: 119.50 Matches: 79  
 Percent Similarity: 35.97% Conservative: 62  
 Best Local Similarity: 20.15% Mismatches: 126  
 Query Match: 6.89% Indels: 125  
 DB: 2 Gaps: 19  
 US-10-617-962-2 (1-1008) x QYRCB8 (1-1485)  
 QY 22 GATGATGAGAGTGAATATCCACCCGTTGAAAGCAATAGCAGAGATATGATGATA 81  
 Db 37 AsnAspAsnGluGlnPheGluGluSerArgLysTyThrIleHisArgGluIleThrArgThr 56  
 QY 82 CTAAAC-----TTTAAAGCAACAGATGAGGCTATACGATCATAT 123  
 Db 57 LeuAsnAlaAsnSerMetLysSerTyArgAsnSerGluGluSerHisLeuGlnAsn--- 75  
 QY 124 GGAATTGAAATATCGAGCTTAAGAAATATATATTAAGCTTACGCTTGGCTGTAAGTGTATT 183  
 Db 76 SerGlnHisTyLys-----TyrAlaSerThrAsnAlaGluVal 88  
 QY 184 CATATGATGATTAATCTTCTGATGACTATATATAAGATTAAGAGACTGTCAGAGAAATT 243  
 Db 89 GlnAsnLysLysGluPheSerAsnAsn-----AsnLysAspAspValGluLeuLys 105  
 QY 244 TATGAGAAATATATGCTTAATCTTCAATCTGCACTATAGTGAAGAAATGTCAGCAAT 303  
 Db 106 ArgProLysAsnTyAsnAsnTyGlyAsnTyLeuAsnThrAsnTyAsnAsnGlyIle 125  
 QY 304 TCTAAAGATATGCAAAATGTTTATTAAGAAATGATGATGATTTGAAGTCAATATCT 363  
 Db 126 AsnTyTyLysAsnValAsnGluLysAsnAsnAsnAlaValLeuSerGlyGlnLysPro 145  
 QY 364 CAATAATTTGGAAT-----GTTCTGAGCTTGAAT- 396  
 Db 146 HisThrIleGlnAsnGluLeuIleValAsnAspHisLysAsnGlyTyThrAsnAsnGlyArg 165  
 QY 397 ---AAACATGAGTGGCTTATTCAGATGAGATTA- 429  
 Db 166 LysLysAsnGluIleSerTyGluAspLysLysMetValGluThrLeuLysAsnGlu 185  
 QY 430 -----TTATTAGCACTATATTTTCTCTGACAGAAATTCACCTGAGGAAAT 480  
 Db 186 LysLysAsnLeuLeuLysIleGluPhePheTyGlyGluSer-----GlnGluAsn 202  
 QY 481 CAACAATCAATGCCCGCAAGATTTTAAATTAATGATTTCTTATTTACTTA- 534  
 Db 203 LysArgIleAsnGln-----LysIleGluAsnPheIlePheThrAsnGluGlu 218  
 QY 534 ----- 534  
 Db 219 LysLeuAsnTySerThrAsnAsnAsnGluTyTyAsnLysTyGluValAsnArgAsnAsn 238  
 QY 535 ---TCTGCTGTAACCTTCACTGGAAGAGATTTTTCAAAATACTTTTAAATGATGATA 591  
 Db 239 AspAsnIleValIleGluAsnIleAsnMetAsnValPheAsnHisAsnGluPheAsnAsnIle 258  
 QY 592 GAGGCTAAATCATTAAGAAATTAATTAATGAGAGAAAAAATCTTTCTTTTGA 651  
 Db 259 LysLysLysAsn-----AsnLeuValSerMetLys----- 268  
 QY 652 CACCGCAGAGATTAACGTGATGAGCAATAGTATTATTTGGCTGACCAAGAGCGCT 711  
 Db 269 -----AspGly- 270  
 QY 712 AATGAGAGTGAAGTTTAAAGAACTTAAATAAATAAATAATCTAGGAATGA- 762

```
Db 271 -----||| |||||:|:|:| |||
QY 763 -----TTTTTAATATGAGAGGGCTGCMAAACAATAT-----AGTTCAATT 807
Db 284 HSHtSphEnAsnGlnAsp-----lGluGluGluTyrArgAspLysAsnSerTyr 301
QY 808 ATAAAAGAGGTACAAAAGGTTACGCTCCACAGACAGCCGAAAAGTATGGTACAGCC 867
Db 302 lLeaAsnAspPhe---LysSerLysIleProGlnTyrAsnSerLysAsnAlaPheSer 320
QY 868 AGTGACGATTAACCTGGAATAATTGCCGAT-----AATTTA 903
Db 321 LysGlySerAsnArgAspIleSerAsnAsnIleAsnAsnAlaGlyValSerAsnMet 340
QY 904 TATAGTGTGAGGCTTAAGCCAAAAGACAGGCTTAACCTTACTCAAAATGATGACAAT 963
Db 341 AsnSerTyrArgMetSerGlyAsnAsnMetTyrArgTyr-----AspThrLysAsn 357
QY 964 ACAATGACGTTTCATATGTTGGAACCTATTATAA 999
Db 358 lMetGlnIleAsnProLysAsnIleHtSgluLys 369
```

## RESULT 4

```
Q8ESK1 PRELIMINARY; PRT; 487 AA.
ID Q8ESK1
AC Q8ESK1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN OrderedLocustNames=O80633;
OS Oceanobacillus theyeensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus theyeensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
CC -1- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;
CC serves to protect cells from the toxic effects of hydrogen
CC peroxide (By similarity).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC EMBL, AF004595; BAC12589.1; -.
DR HSPSP; P04040; 1DGF.
DR GO; GO:0004096; F:catalase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR002226; Catalase.
DR Pfam; PF00199; Catalase; 1.
DR PRINTS; PR00067; Catalase.
DR ProDom; PD000510; Catalase; 1.
DR PROSITE; PS00437; CATALASE_1; UNKNOWN_1.
DR PROSITE; PS00438; CATALASE_2; 1.
KW Complete proteome; Heme; Hydrogen peroxide; Iron; Oxidoreductase;
KW Peroxidase.
SQ SEQUENCE 487 AA; 55391 MW; 841B2B318C7ACAB8 CRC64;
```

## Alignment Scores:

```
Pred. No.: 1.38
Score: 116.50
Percent Similarity: 35.11%
Best Local Similarity: 22.57%
Query Match: 6.71%
DB: 2 Gaps: 18
```

## US-10-617-962-2 (1-1008) x Q8ESK1 (1-487)

```
QY 160 TACGCTTTGGCTGTAGTGTATTCATATATGATTAATCTTCTGATGAC----- 210
Db 168 TrpSerLeuSerProGluSerLeuHtIleGlnValThrTyrLeuHtSerAspArgGlyLeu 187
QY 211 -----TATTATAGAAATAA 225
Db 188 ProAlaThrLeuArgHtMetAsnGlyTyrGlySerHtIleThrPheLysTyrValAsnAla 207
QY 226 GAGACTGCTGAGAGAAATTTATCAAGAAATAT-----ATGCTAAT 264
Db 208 Glu---GlyGluAlaPheTrpValLysTyrHtSphHtSerGluGlnGlyLysAsn 226
QY 265 CTTCATCTGCACTA-----TTAGGTGAAAATGGTGATCAAAATTTCTAAAGAT 312
Db 227 LeuAspProAlaLeuAlaAspLysIleAlaGlyGluAsnProAspTyrHtIleThrGluAsp 246
QY 313 ATGCAAAATGGTTTTTATTAAGATGAAGTGAATTTTGAGGTCAATATCTCAAAACATT 372
Db 247 LeuPheAsnAlaIleGlu-----GluGlyAspHisPro-----Ala 258
QY 373 TGAATGTTCCGAGCTTGAAAATTAACCATTAAGTGCTTATTCAGATGAGATTAATTA 432
Db 259 TrpLys----- 260
QY 433 TTAGCATATATTTTCTCTGTACAGAAATTCACGTGAGAGAAATCAACATCAAAAT 492
Db 261 -----LeuTyr-----ValGlnIleMetProTyrGluAsp----- 270
QY 493 GCCGCAAGATTTTAAATTAATTAATGATTTCTATTACTTATCTGCTGTAACTTCACGTG 552
Db 271 --AlaLysThrTyrLysTrpAspProPheAspValThr----- 282
QY 553 GGAAGAGAGATTTTTCAAAATACTTACAAATGATGATTAAGGCT----- 597
Db 283 -----LysIleTrpSerLysLysAspTyrProArgIleGluValGlyArgMetValLeu 300
QY 598 --AAATCATTAAGAAATTAAT-----ATTGAGAGAAAAAACTTTCTAAACCTTTC 645
Db 301 AsparGAsnProGluAsnTyrPheAlaGluValGluGlnAlaIaPheSerProGlyGln 320
QY 646 TTTCGACCA-----CCGCAAGATTAACGTATGGCAGAAATAGTTATTGG 690
Db 321 PheValProGlyIleGluAlaSerProAspLysMetLeuGlnGlyArg-----Leu 337
QY 691 GCTGACCAACAGAAAGCCTAAATGAGAGATGAGTTT----- 729
Db 338 PheGlyTyrSerAspAlaHtIleArgTyrArgLeuGlyAlaAsnHtSerIleProVal 357
QY 720 -----AAAGACTTAAATAATTAACAATCTAGGAATGATTTTCTAATATGGAAGG 780
Db 358 AsnArgProLysAsnGluValAsnAsnTyrGlnArgAspGlyPheMetSerValAspGly 377
QY 781 GCTGCAAAACAAAAGTATGTTCAATTATTAAGAAGGTACAAAAGGTTAAGCTCCACAG 840
Db 378 -----AsnGlyGlyAsnLysProAsn 384
QY 841 ACAGCAGGAAAAAGTAT--GTTACAGCCAGTGCAGTAACCTCGAATAATTTGCCGAT 897
Db 385 TyrGluProAsnSerValAsnGlyProThrGluAspSerGlyAlaLysIleAsnProPhe 404
QY 898 AATTATATATAGTGTAGGCTTAAGCCAAAAGACAGGCTTAACCTTACTCAAAATGAT 954
Db 405 GuileTyr-----GlyGlnAlaAspSerValThrTyrAspSerAspAsp 419
```

## RESULT 5

```
TIG_PROMP STANDARD; PRT; 473 AA.
ID TIG_PROMP
AC Q7UZX8;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
```

DE Trigger factor (TF).  
 GN Name=Trig; OrderedLocNames=PMM1655;  
 OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).  
 CC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
 CC Prochlorococcus.  
 CC NCBI\_TaxID=59919;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;  
 RA Rocap G., Latimer F.W., Lamerdin J.E., Malfatti S., Chain P.,  
 RA Ahlgren N.A., Arcillano A., Coleman M., Hauser L., Hess W.R.,  
 RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,  
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
 RA Webb E.A., Zinser E.R., Chisholm S.W.;  
 RA "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
 RT niche differentiation".  
 RL Nature 424:1042-1047(2003).  
 CC -I- FUNCTION: Involved in protein export. Acts as a chaperone by  
 CC maintaining the newly synthesized protein in an open conformation  
 CC (By similarity).  
 CC -I- SIMILARITY: Belongs to the FKBP-type Ppiase family. Trg subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL: BX572094; CAE20114.1; -  
 DR HAMAP: MF\_00303; -; 1.  
 DR InterPro: IPR001179; FKBP\_Ppiase.  
 DR InterPro: IPR008880; Trigger\_C.  
 DR InterPro: IPR008881; Trigger\_N.  
 DR InterPro: IPR005215; Trig\_fac.  
 DR Pfam: PF00254; FKBP\_C; 1.  
 DR Pfam: PF05698; Trigger\_C; 1.  
 DR Pfam: PF05697; Trigger\_N; 1.  
 DR TIGRFAMs: TIGR00115; Cig; 1.  
 DR TIGRFAMs: TIGR00115; Cig; 1.  
 DR PROSITE: PS50059; FKBP\_PPIASE; 1.  
 DR Cell division; Chaperone; Complete proteome; Isomerase; Rotamase.  
 FT DOMAIN 174 261 Ppiase, FKBP-type.  
 SQ SEQUENCE 473 AA; 53533 MW; 31142AP2CB4E9AED CRC64;  
 Alignment Scores:  
 Pred. No.: 1.65 Length: 473  
 Score: 115.50 Matches: 76  
 Percent Similarity: 39.88% Conservatve: 52  
 Best Local Similarity: 23.68% Mismatches: 146  
 Query Match: 6.66% Indels: 47  
 DB: 1 Gaps: 15  
 US-10-617-962-2 (1-1008) x Trg\_PROMP (1-473)  
 QY 43 CCCGTTGAAGCAATAGCA-----GAGATATAGTACTATTAACCTT----- 90  
 DB 165 ProValaamamaryproalalyaleuglyaspilleaia--ValValSerPheylsegly 183  
 QY 91 -----AAGCAACAGATGAGGTCATACATCAATATGAAATTGAA 132  
 DB 184 ValTylrlyasapserlyaleuglyaleuglyaleuglyaleuglyaleuglyaleugly 203  
 QY 133 TATCGAGCTAAGAAATATATATTAAGCTTACGCTTGGCTGTAAGTGTATTCAT----- 186  
 DB 204 Leuglulysaenlymetilleprogllyphevalgluglylilevalgllymetlyalleasp 223  
 QY 187 -----AAGTATCTAACTTCTGATGACTAT-----TATAAGAAATAAGAGACT 231  
 DB 224 AspaenlysthlLeuthrleuargphepogluaspilyserHlsigluspserarggly 243  
 QY 232 GCTGAGAGATTTATCAAGAAATATATGTCTATCTTTCATCTGCACTATTAGTGAAAT 291

DB 244 LysglulalallepaePllSerleuylsaspLeuylsGluLeuProgluLeu 263  
 QY 292 GGTGATCAAAATTTCTAAAGATATATGCAAAATGTTTATTAAGAAATGAACTG-----GAT 345  
 DB 264 AsmpasapPhealalyglinserclyasnllyasapserleuylsGluLeuylsasp 283  
 QY 346 TTTGAGGTCATATATCTCAAAACATTTGAAATGTTCTGAGCTTGAATAATTAACCATTTG 405  
 DB 284 lileglulysgluleuylsGluAenPheapsernthrgln-----LysaenlleysVal 301  
 QY 406 AGTGCTTATCAAT-----GACGATTAATTTATGACATATATTTTCTCTGACAG 459  
 DB 302 GlulaleuMetaspalaleuSerlysegluleuaspalagluilleProlysalMetille 321  
 QY 460 GAAATTCACCTGAGAGAAATCAACAATCAATATGCGGCAAGATTATTTTAAATTA----- 513  
 DB 322 Asplleugluvalargasasnlileglulnthralaglnargpnealaglndlymet 341  
 QY 514 ---ATGATTTCTTATTTACCTTATCTGCTGTAACCTGCTGGA-----AGAGAG 561  
 DB 342 AsplleuSerthrPheThrProgluleuVallyserleuAlagluSerThrArgPro 361  
 QY 562 ATTTTCAAAAACCTTTTACAAATGATTAAGGCTTAATCATTTAGAAATTTATTTAG 621  
 DB 362 GlulalagluysaenlyaleuglnargasnlleuylsAlaleuSer-----Glu 378  
 QY 622 AGAAAAAATCTTCTTCAAACTTTCTTTCGACACGACGAGATTACTGATGAGAGAAATA 681  
 DB 379 ArgglulyslethrValaspasnlglulileaspglulysMetlyleGlu----- 395  
 QY 682 GGTATTTGGCTGAGCAACAGAGAGCGCTTAATGAGATGATGATTTTAAACCTTAA 741  
 DB 396 ---TyrGlusapglulileserlyserProlys---Glnlleaspilleglulysleuyls 413  
 QY 742 AATAACAAATTTAGAGATGATTTTCTTATATGCAAGGCGCTGCAAAACAAAGTATAGT 801  
 DB 414 ---AspValValArgasaspLeuLeuGlulysleuilethrTrpleuglulasn 432  
 QY 802 TCATTTATTAAGAGGTCACAA-----AAGGTTAACCTCTCA 837  
 DB 432 SeralaleuylsGluileaspnglulysAlathrlyleuThrlyThrThrlyls 452  
 QY 838 CAGACAGAGAGAGAAAGATTTGTCACGACGTCAGCACTGCAAAATTTGCCGAT 897  
 DB 453 AlathrThrlylsGlyVallyThrlyserlyProlysValasnllyleuglulys 472  
 QY 898 AAT 900  
 DB 473 Aen 473  
 RESULT 6  
 Q81226 PRELIMINARY; PRT; 434 AA.  
 AC Q81226;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Tatp-Like deoxyribonuclease, putative.  
 GN Name=PFA0580C; Synonyms=MALIP3\_01;  
 OS Plasmodium falciparum (isolate 3D7).  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CC NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Bettman M., Churher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corson C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Fellwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagsels K., James K.D., Johnson D., Kerhornou A.,



```

Db      126  GluYrThrAlaValIleSerAsp-----LyHieGlyLeuVal 139
Qy      346  TTGAGAGTCATATCCCAAAACCTTGGATGTCCTGAGCTGAAATAAACAATG 405
Db      140  AsnAspLysGlnIYrSerAsnAsn-----HisSerLeuAsnProHe 153
Qy      406  AGTCCTTATTCAGATGACGATAAATTA---TTAGCACTATATTTTCTGTCAGAGAA 462
Db      154  AsnLeuPheGluAspAspGluIleAsnIYrSerIleYrYr----- 168
Qy      463  ATTCACCTGAGGAAATCAACATCAATGACCCGACAGATTTTAAATTAATGATTTC 522
Db      169  -----LysGluYrSerGluAsnGlnIurHeuThrArgIprGlnValLeuGlyIuHe 186
Qy      523  TTATTTACCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 582
Db      187  GlyGluAsnGlnIurPheAsnIleProSerIleGluAlaLeuAsnGlyIleAsnAsn 206
Qy      583  -----AATGATTAAGAGGCTAAATCATTAAGATTAATTAATTAAGAGGAA 627
Db      207  IleIleGlySerAlaAspAsnThrIYrValThrIYrLeuThrAspThrIYrSerThr 226
Qy      628  AAACCTTTTAAACCTTTCTTTCGACACGACGAGATTAACCTGAGCAGAAATGATT 687
Db      227  ValLeuAsnLys-----AspThrIleIleAspGlu 236
Qy      688  TTGCGTCGACCAACAGAACCCCTTAATGAGAGTGAATTTTAAACCTTAATAATAC 747
Db      237  ValThrGlyProPheLeuValAlaLys-----Asn 246
Qy      748  AAATAGTAGATGATTTTAAATATGAGAGGCTGCAAAACAATAATAGTCAATT 807
Db      247  GluGlnGluGlnIleGlyPheSerIleYrAsnGlu---IleMetLysIYrAspPheGln 265
Qy      808  ATAAAGAGGATCAAAAAGGTTACCT-----CCACAGACAGCAGCAGCAAGATT 858
Db      266  Ile-----AsnAlaIleIleIleProAspLeuIYrThrGlnLeuIle 279
Qy      859  GGTACAGCCAGTCGACGATACCTGCAAAATTTCCGAATTAATTATATGTCGAGCTA 918
Db      280  -----SerLeuAlaIYrLeuAspSerAsnLeuIYrSerThrAsnLeu 293
Qy      919  AGCCAAAAGACAGGCTA---ACCTTACTCAAAATGATGATGCAATTAATGACGCTT 975
Db      294  ThrThrThrAsnIYrThrArgSerPheValAlaGluAsnThrSerAsnLysLeu---Val 312
Qy      976  CATAGTGT 984
Db      313  SerSerVal 315

RESULT 8
Q81SF7 PRELIMINARY; PRT; 908 AA.
AC Q81SF7;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PFL1255C;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=36329;
RX SEQUENCE FROM N.A.
RA MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Sun B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Hult D., Mather M.W., Valdivia A.B.,
RA Martin D.M., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

```

```

RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511 (2002).
RN (2)
RP SEQUENCE FROM N.A.
RA Hyman R.W., Fung E., Conway A., Kurd O., Mao J., Miranda M.,
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;
RL Submitted (JAN-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB014848; AAN36337.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 908 AA; 108485 MW; 48FPE8872F7C670 CRC64;

Alignment Scores:
Pred. No.: 2.67 Length: 908
Score: 113.00 Matches: 74
Percent Similarity: 36.66% Conservative: 51
Best Local Similarity: 21.70% Mismatches: 112
Query Match: 6.51% Indels: 104
DB: Gaps: 18

US-10-617-962-2 (1-1008) x Q81SF7 (1-908)
Qy      127  ATGAATATCGAGCTAAGAAATATTAATTAAGCTTACGCTTGGCTGTAAGTGTATTCAT 186
Db      168  LeuYrIYrIleProLeuLysIleIleYrThrGln-----SerThrLeuHe 183
Qy      187  AATGATCTAAACCTTCTGATGACTTATTAAGAAATTAAGAGACTGAGAGAAATTTAT 246
Db      184  AspIleSerAsnLeuLysAsnGluIleYrLys----- 194
Qy      247  CAAGATATATGCTAATCTTTCATCTGACATTAATAGTGAATATGTCATCA----- 300
Db      195  -----TyrIleAsnGluValIleLeuMetSerGlnGlyAspGlnAsnIle 209
Qy      301  ATTCTAAAGATATGCAAAATGATTTTATTAAGATGCAATTTTGAAGTCAATAT 360
Db      210  LeuSerLysValGluIleIleLysAsnIleLysAsnAsnIleAsnSerGluYrSerYr 229
Qy      361  CCTCAAAACATTTGGAATGTTCTGAGCTTGAATAAATCAATGAGTCTTATTCAGAT 420
Db      230  AspHisLeuValThrSerLysPheIleAspThrGlnAsn---AspIleYrSerAsp 248
Qy      421  GACGATTAATTAATGCA-----CTATATTTTCTCTGACAGAAATTCACGTGAG 474
Db      249  AspHisIleIleValGlnAspAsnAsnCySerPheThrThrThrGluYrIleAsnIleGlu 268
Qy      475  GAAATCAACATCAATGACCCGCAAGATTTTAAATTA----- 513
Db      269  AsnAsnLeuLys---AsnIleTrpAsnLeuLeuSerValHisIleHisAsnAsnIleCys 287
Qy      514  ---ATTGATTTCTTATTAACCTTATCTGCTGTAACCTTCACTGGAGAGAGATTTTCA 570
Db      288  AsnValAspIleLysPheThr-----GlyLys-----LeuSer 298
Qy      571  AAAAATCTTTCATATGATTTGAGAGCTTAATCATTA-----GAGAAAT 612
Db      299  ProArgPheYrGlnIYrValAsnMetLysGluLeuLysAspIleProAsnIleGluLys 318
Qy      613  TATATTTGAGAGAAAAAATCTTCAACCTTCTTTCGACACGCGACAGATTAACCTGAT 672
Db      319  Tyr---GluAsnIleAsnPheSerSerLeuPheYrIYrAsnPro---LysIleAsnAsp 336
Qy      673  GCGAGATATGCTTATTTGGCTGACCAACAGAGAGCCTTAATGAGAGTGAAGT----- 726
Db      337  GluYrIleLysAsnIleValAsnGlnThr---IleProHisTrpSerValCysGluGlu 355
Qy      727  -----TTTAAAGACTTAATAAATTAACAAATCTAGCAAT----- 759
Db      356  LysLysLysIleGlnIleAspYrGluLysGluIleLysMetAspLysAsnAsnAsp 375
Qy      760  -----GATTTTCTAAT 771

```





```

Qy 778 -----GGGGCTGCACAAACAAAGTATGTTCAATT 807
Db 977 ArgilleusleupheapgluValgluLeuasnHisValylshisSerAsnlyMetlle 996
Qy 808 ATAAAGAGGTACAA-----AAGGGTAAAGCTCCACAGACGACGCAAAAGTATGGT 861
Db 997 TyrlsEglnValTyrlsPheVallyGlyAsnThr---GluSerValSerPheGluIleasp 1015
Qy 862 ACACCCAGCTGCAGTAACTCGAAAATG----- 891
Db 1016 SerlleasnlysglyTyrlslelyMetlyslslyAsnTyrlslyGluHisMet 1035
Qy 892 -----CCGAATTAATTATATAGTGTGAGAGCTAAGCCAAAGACAGGTA 936
Db 1036 AsnlysbAsnAsnlysbAspAsnAsnAsnAsnAsnAsnSerAsnlysbAspAsnHisIle 1055
Qy 937 ACCTTACTCAAAATGAT----- 954
Db 1056 AsnIleasnMetAsnAspAsnHisIleArgAsnTyrlsAsnAspIleAsnleuglyProAsnSer 1075
Qy 955 ACTGACATACATGACGCTTCATGCTTGAATCTATAT 996
Db 1076 ThrspAspSerProThrValSerSerleuglyAsnGluTyrls 1089

RESULT 10
Q7RT20 PRELIMINARY; PRT; 995 AA.
AC Q7RT20;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=FY00181;
OS Plasmodium yoelii yoelii;
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Koof J.T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoat A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL: AABL0100048; EAA21179.1;
DR InterPro; IPR001862; MAC perforin.
DR Pfam; PF01823; MACPF; 1.
KW Hypothetical protein.
SQ SEQUENCE 995 AA; 114106 MW; CC0889FD84EB63EA CRC64;

Alignment Scores:
Pred. No.: 3.85 Length: 995
Score: 111.00 Matches: 68
Percent Similarity: 37.17% Conservative: 45
Best Local Similarity: 22.37% Mismatches: 97
Query Match: 6.40% Indels: 94
Db: 2 Gaps: 15

US-10-617-962-2 (1-1008) x Q7RT20 (1-995)
Qy 187 AATGATCTAAACTCCCTGATGATGATATAT-----AAGAAAT 222

```

```

Db 615 AsnMetlyThrIlePheAspPheTyrlsPheIlelyMetglyLeuSerVallyArgAsn 634
Qy 223 AAAGAGCTGCTGAGAAATTTATCAAGAAATATATGCTATCTTCACTGACCTATTA 282
Db 635 LysglValgluTyrlsPheIleAsnlyslleIlelySerlyThrValSerlleugly 654
Qy 283 GGTGAATAATGCTGATCAAAATTTCTAAAGATATGCAATAGTGTTTTAAAGATGAAC 342
Db 655 GlyHisProglyLeuAsnIleAspAsp-----SerSerPhePheGluTyrls 671
Qy 343 GATTTGAAGCTCAATATCTCAAAACATTTGAAATGTTCTGACCTGAAAATPAA 399
Db 672 AspSerlleAspLylys-----AsnSerMetProIleArgThrlyAsn 685
Qy 400 -----CCATTGAGTGTATTATGATGACGATTAATTTATAGACATATATTTTCTCTGA 456
Db 686 LeuProPheSerPhePheMetAspAspProAsnMetllelysalatyrls----- 701
Qy 457 CAGGAATTCACCTGAGAGAAATCAACAATCAATGATGCGCAAGATTTTAAATTAAT 516
Db 702 -----AsnAspIleuMetPheTyrlsGlyLeu 710
Qy 517 GATTTCTATTATTAACCTTATCTGTAACCTGACGAGAGAGATTTTCAAAAAC 576
Db 711 -----ThyTyrlsValgluIlePheAspGlnAsn 720
Qy 577 TTTTACAATGATTAAGAGCTTAATCATTTAGATTAATTTATGAGAGAAAAAACCTTCT 636
Db 721 GlnTyrlsAsnAsnAsnGluIle-----SerllelyAsnTyrlsGlu-----LysSerlleGln 738
Qy 637 AAACCTTCTTTCAGACACGCGAGATTAACCTGATGCGAGATAGGT-----TATTTG 690
Db 739 LysIleTyrlsIsglySerProglyLeuLeuThrCysProIleIsglySerThrIleLeu 758
Qy 691 GCTGACCAACAGAGAGCGCTTAATGAGAGCTGATTT----- 729
Db 759 MetglyPheSer-----LeuAsnleuAspPheTyrlsAsnGlnAsnleuAsn 774
Qy 730 -----AAGAACTTTAAATCAACAATCTAGAAATGATTT 765
Db 775 GluIleIleIleGlyIleAsnAlaCysgluIleMetlysglySerCysSerGlyAsnIlePhe 794
Qy 766 TCTAT-----ATGAGAGGGGCTGCAAAACAAGTATATGTTCAATT 807
Db 795 ThrAsnlyTyrlsSerAspIleArgIleTyrlsIleuTyrlsSerGlyIleuTyrlsPhe 814
Qy 808 ATAAAGAG-----GTACAAAGAGGTAAAGCTCCACAGACAGACGCG----- 849
Db 815 IlelysglnValgluGlnAsnGluSerThrlystrThrAlaThrCysProGluAsp 834
Qy 850 -----AAAATATTGTTGATCAGCGACGAGCTTAAC 879
Db 835 SerValIleleuPheglyPheAlaLeuMetlysglyIleIleGlyArgSerSerAlaAsnThr 854
Qy 880 CTGGAATAATGCGGAATTAATTATATGATGAGGCTAAGCAAAAGACAGGTAAC 939
Db 885 ValAsp-----LeuTyrlsProCysArgThrGlyGlnAsnSerCysSerAla 869
Qy 940 TTTACTCAAAAT 951
Db 870 ValleuglnAsn 873

RESULT 11
Q8IHV6 PRELIMINARY; PRT; 1794 AA.
AC Q8IHV6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF11_0420;
OS Plasmodium falciparum (isolate 3D7).

```



```

Query Match: 6.34% Indels: 102
DB: 2 Gaps: 13
US-10-617-962-2 (1-1008) x Q8II04 (1-3468)

QY 67 GATATAGTACGTATATCTTAATCTTTAAGCAAAACAGATGAGGCTCATACGATCATATGGA 126
   |||||
Db 872 AsnValIleHisAsnLysAsnPhelLysIleAsnGluIleAspHisPhelLysGluLysAsn 891

QY 127 ---ATTGAATATCGATGAGAAATAATATATTAGCTTACGCTTGAGTGAAGTGT--- 180
   ||| ||||| |||||
Db 892 ThrLeuAsnTyrGluSerLysAsnIleIleAsnThrCysLysAspAsnLeuGluLysAsp 911

QY 181 -----ATTCAATATGATATCTTAACCTTCTGATGAC----- 210
   ||||| ||||| ||||| |||||
Db 912 GluAsnAsnIleAsnAsnIleAsnGluLeuLysAsnAsnSerLeuGlnPheAspLysAsn 931

QY 211 ---TATTATTAAGATTAAGAGACTGCTGAGAGAAATTTATCAAGAAATAT--- 255
   ||||| ||||| ||||| |||||
Db 932 IleLeuLeuLysAsnThrMetSerLeuGlnLysAspTyrSerAsnIleLysAsnArgLys 951

QY 256 -----ATGCTTAATCTT 267
   ||||| ||||| ||||| |||||
Db 952 AlaAsnThrSerAsnIleAspAsnIleAsnLysTyrThrSerAsnAsnIleSerAsnLys 971

QY 268 TCATCTGCACTATTAGTGAATAATGTCATCAAAATTTCT-----AAAGATATGCA 318
   ||||| ||||| ||||| |||||
Db 972 SerAspIlePheIleAspAsnAsnLysSerLysSerTyrAsnLysThrAspIleIle 991

QY 319 AATGCTTTTATTAAGATGAAGCTGATTTGAAGTCAATATCTCAAAACATTTGGAAT 378
   ||||| ||||| ||||| |||||
Db 992 AsnIlePheSerLysAsnLysSerAsnAspGluAspThrPheSerLysCysPheThrTyr 1011

QY 379 GTTCTGAGCTTGAATAAACCATGAGTGTATTTCAGATGAGATTAATATTATACA 438
   ||||| ||||| ||||| |||||
Db 1012 LysGluHisLeuSerAsn-----TyrAsnLysAspGluAsnLeuPhe 1026

QY 439 CTATATTTTCTCTGTACAGAAATTCACCTGAGGAAAGAAATCAATCAATCCGCCA 498
   ||||| ||||| ||||| |||||
Db 1027 ThrThrPheSerAsnAlaPheHisIleProLysIleAsnAsnAsnIleLysSer----- 1044

QY 499 AGATTTTAAATTAATGATTTCTTATTTACCTTATCTGCTGAATCTCACTGAGGAAG 558
   ||||| ||||| ||||| |||||
Db 1045 -----ThrHisAsnAspIleLeuAsnIleSerAsn 1054

QY 559 AGCATTTTCAAAAACTTTTACATGATTAAGGCTTAATCTTGAAGATTAATATT 618
   ||||| ||||| ||||| |||||
Db 1055 LysTyrTyrAsnLysAsn-----AsnThrThrAsnLysThrAsnIleAspIlePheGln 1072

QY 619 GAGAGAAAAAACTTTCTTAACCTTTCTTGACACCGCAGAGATTACCTGATGCGACA 678
   ||||| ||||| ||||| |||||
Db 1073 AsnLysAsnSerLeuAspMetPro----- 1080

QY 679 ATAGGTTATTTGGCTGACCAAGAACGCTTAATGAGAGATGATTTT----- 729
   ||||| ||||| ||||| |||||
Db 1081 -----GluThrAsnLeuIleLysGluLysGlnGlnPheIleLysThr 1094

QY 730 -----AAAGAACTTAATAATCAAAATCTAGAGATGATTTTCTAAT 771
   ||||| ||||| ||||| |||||
Db 1095 AspMetSerSerTyrAspAsnSerIleLysAsnAspAsnSerArgAspIleArgGluAsn 1114

QY 772 ATGGAAGGGGCTGCAAAACAAAGATAGTTCATTTATTAAGAGAGTCAAAAGGCTAC 831
   ||||| ||||| ||||| |||||
Db 1115 IleAspCysSerValLysAsnGluTyrGlnSerPheAsnLysAsnMetSerAsnLysAsn 1134

QY 832 GCTCCACAG----- 840
   ||||| ||||| ||||| |||||
Db 1135 LeuGlySerMetAsnValSerProCysAsnHisMetTyrAsnGluAsnAsnAspIleIle 1154

QY 841 ACAGACGGAAGAAAGTATTGGTACAGCCAGTGGCAGTAACTTGGAAAA-----TTGCCG 894
   ||||| ||||| ||||| |||||
Db 1155 AsnIleSerLysTyrIleGluHisArgIleArgLysAsnAlaGluGluAsnThrMetLys 1174

```

```

QY 895 AATAATTATATAGTGTGAGGCTAGCCAAAAACAGAGGTTACTTACTCAAAATGAT 954
   ||||| ||||| ||||| |||||
Db 1175 AsnAsnIle---AsnGluAspThrSerSerLysAspPheAsnCysPheSerAsnAsnGlu 1193

QY 955 ACTGACATCAATGACGCTTCATAGTGT 984
   ||||| ||||| ||||| |||||
Db 1194 LysLysLysAsnPheThrThrAsnAsnIle 1203

RESULT 13
Y028 BORBU STANDARD; PRT; 414 AA.
ID Y028 BORBU
AC OE0959;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hypothetical ANK-repeat protein BB28.
GN OrderedLocustNames=BB28;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Caesteens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Kerchum K.A., Dodson K.J., Hickey E.K.,
RA Gwynn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kierlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., Van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Matthey L., McDonald L.A.,
RA Artchak P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586 (1997).
CC -1- SIMILARITY: Contains 2 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE000792; AAC6335.1; -.
CC PIR: H70219; H70219.
CC TIGR: BB28; -.
CC InterPro: IPR002110; ANK.
CC Pfam: PF00023; ANK, 2.
CC SMART: SM00248; ANK, 2.
CC PROSITE: PS50088; ANK_REPEAT, 1.
CC PROSITE: PS50297; ANK_REPEAT_REGION, 1.
CC ANK repeat; Complete proteome; Hypothetical protein; Plaemid; Repeat.
FT REPEAT 326..355 ANK 1.
FT REPEAT 359..389 ANK 2.
SQ SEQUENCE 414 AA; 49379 MW; B6ACASD8PA2B6680 CRC64;

Alignment Scores:
Pred. No.: 4.82 Length: 414
Score: 109.50 Matches: 72
Percent Similarity: 36.02% Conservative: 44
Best Local Similarity: 22.36% Mismatches: 98
Query Match: 6.31% Indels: 108
DB: 1 Gaps: 14

US-10-617-962-2 (1-1008) x Y028_BORBU (1-414)

QY 10 CAATTAACCTGATGATAGAGAGATATCCACCCGTTGAAGCAATATGACAGAGAT 69
   ||||| ||||| ||||| |||||
Db 178 GluIleSerIleAspGluAsnGlyAsnIleIleProValIleLysAsnThrAsnLysGln 197

```

```
QY 70 ATAGTACGTATTA-----AACTTAAGCAACAGATGAGGTATACAGCA 117
Db 198 ILeThrgluLeuLeuPheThyAspGluAsnTyAsnLeuLeuProPheLysPheGlu 217
QY 118 TCATATGAAATGAAATAT-----CGAGCTAAGAAATA 150
Db 218 SerTyLysValArgPheAspLeuLeuProGluAsnLysAsnLeuAsnPhelGluLeu 237
QY 151 ATATTAGCTTACGCTTGGCTGTAAGTGTATTCATTAATGATCTAACTTCGATGAC 210
Db 238 ILeAsnValTyTyLeuAspGluLysAsnLeuLeuThr-----ProLeuGlu 253
QY 211 TATTATTAAGATAAAGAGCTGCTGAGAGATTTTCAAGATATATGCTATATCTTTCA 270
Db 254 TyTyTyLysAsnAsnLeuAspMetSerProTyTyTyLeuAsp----- 267
QY 271 TCTGACATATTAGGAAATAATGATGATCAAAATTTCTAAAGATATGCAAAATGTTTAT 330
Db 268 -----LeuGluGluAsnLysAspAsp-PheLeuLys----- 277
QY 331 AAGAAATGAACTGATTTTGAAGTCAATATCTCAAAACATTGGAATGTCGAGCTT 390
Db 278 -----AlaIleLysIleLysLysGluTyGluTyIleGluLys 291
QY 391 GAAATTAACCATTTGAGTCTTATTCAGATGACGATTAATTTTACATATTTTTC 450
Db 291 sLys--LysGluLeuGluAsnLeuThrgluAsnAspLysLeu-----Asp 305
QY 451 TCTGACAGGAATTCACCTGAGGAAATAATCAACATCAAAATGCGCAAGATTTTAA 510
Db 306 AspPheLysGluPheLeuLeuLysAsnAsn----- 316
QY 511 TTAATGATTTCTTATTTTACCTTATCTGCTGCTTACTTCACTGGAAGAGATTTTCA 570
Db 317 -----IlePheSerLeuAsnThrxIlePheSerAsnGlyAsnProIlePheThr 332
QY 571 AAAAATTTCATATGATGATTTAGAGCTTAATCATTTAGAGATTTATTTAGAGAAAA 630
Db 333 -----TyraIleAsnValLysAlaLysSerIleIleAsnTyLeuLeuThrxLysGlu 350
QY 631 CTTTCTAAACCTTTTTCGACACCGAGAGATTACCTGAGGCAATAGCTTATTG 690
Db 351 PheAsn----- 352
QY 691 GCTGACCAACAGAACGCTTAATATGAGAGTGTATTTAAAGATTAAATTAACAA 750
Db 353 -----IleAsnLeuThr 356
QY 751 TCTAGGAATGATTTTCTAATATGAGAGGCTGCAAAACAAAGTATGCT--TCATTT 807
Db 357 AsnGluAsnSerGluThraIleAsnLysSerAlaIleIleGluLysTyAspLeuAsnPh 376
QY 808 ATAAAGAG--GTACAAAGGGTAAACGCTCCACAGACAGCAGCAAAAGTATGTTGTA 864
Db 377 ILeLysSerLeuIleGluLysGlyAlaAsnPro-----AsnIle 389
QY 865 GCCAGTGTACGATTAACGTAAGAAATTTGCCGAATTAATTTATAGTGTGAGCTTAAG 924
Db 390 ArgAspGlyAspAsn-----LysLeuProIleAsp--TySerAspLysThrSerGlu 406

RESULT 14
Q7RNS5 PRELIMINARY; PRT; 1491 AA.
AC 07RNS5;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY01369;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=73239;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlson J.M., Angluoli S.V., Suh B.B., Koof T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.B., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
RA van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carnocci D.J.,
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL, AABU0100539; EAA21375.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1491 AA; 175817 MW; 7FA33125E6B1F26E CRC64;

Alignment Scores:
Pred. No.: 5.15 Length: 1491
Score: 109.50 Matches: 76
Percent Similarity: 41.96% Conservative: 57
Best Local Similarity: 23.97% Mismatches: 104
Query Match: 6.31% Indels: 80
DB: Gaps: 17

US-10-617-962-2 (1-1008) x Q7RNS5 (1-1491)
QY 187 AATGATCTAAACTTCTGATGATCTATTAAG-----AATAAAG 228
Db 11 SerLeSerLysTyTyGluAspAspTyTyAsnLysGluLeuThrgluAsnGlu 30
QY 229 ACTGCTGAGAAATTTATCAAGAA-----TATATGCTAATCTT--TCATCTGCA 276
Db 31 AsnAlaLys-----TyTyGluAspAsnAlaTyTyGluThraAsnIlePheGluAsnSer 48
QY 277 CTATTAGGTGA--AATGATGATCAATTTCTAAAGTATGCAAAATGCT--TTTAT 330
Db 49 ILeHisProGluIleLysAsnAsnLysPheProLeuAspTyTyTyAsnSerGluGlyTy 68
QY 331 AAGATGAAGTGAATTT-----GAAGCTAATATCTCAAAACATTGGAATGTTCT 384
Db 69 LysAsnGluAlaAspTyMetGluLysAsnLysTyTyGluSerCysIleAspAsnLysGly 88
QY 385 GAGCTTGAATAATTAACCA-----TTGAGTGCTTATTCAGATGACGATTAATTA 432
Db 89 AsnSerMetAsnLysProAsnMetGluThrPheGluAsnIleAsnAspHisGluLys-- 107
QY 433 TTAGCATATATTTTCTCTGTACAGAAATTTCCACGTGAGGAAATCAACATCAAT 492
Db 108 -----IleGluGluAsnLysIleLysAsn 115
QY 493 GCCCGAGATTTTAAATTAATGATTTCTTATTACCTATCTGCTGAATCTCACTG 552
Db 116 GlyAspArg-----SerIleSerGluSerIleAsnTy 126
QY 553 GGAAGAGATTTTTCAAAAAATTTCACATGATTAAGGCTAAATCTTAAG-- 609
Db 127 AsnLysGluAsnThrgluMetAsnLeuIleAsnIleLysGluAsnLysThrgluGlu 146
QY 610 -----ATTATATTGAGAGAAAAAATTT-----CTAAACCTTTCTTCGACCA 657
Db 147 LysThraAsnPhelGlyAsp-AsnLysAsnPhaAsnLeuLysLysLysLeuThraAs 166
QY 658 CAGAGATTACCTGATGAGCAATAGTATTTGCTGAGCAACAGAGCGGCTTAATG 717
Db 166 pGluAspAspLeuIleAspAspLeuLeuIleTyTyAsnAsnAsnGluGluTyGlyAsnTh 186
```



Db 1047 LysGlnGluAsnIleLysIleHisMetGluProSerHisIleSerAsnIle 1063

Search completed: November 21, 2004, 08:54:00  
Job time : 161.5 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	1008	100.0	1008	16	US-10-617-962-2	Sequence 2, Appl1
2	663.8	65.9	1107	16	US-10-617-962-1	Sequence 1, Appl1
3	663.8	65.9	1205	16	US-10-617-962-5	Sequence 5, Appl1
4	50.6	5.0	188971	16	US-10-235-192A-7	Sequence 27, Appl1
5	49.8	4.9	11394	15	US-10-240-453-95	Sequence 95, Appl1
6	48.8	4.8	1866	15	US-10-349-680-73	Sequence 73, Appl1
7	48.2	4.8	2767	15	US-10-301-533-23	Sequence 23, Appl1
8	47.6	4.7	16217	15	US-10-311-445-597	Sequence 597, Appl1
9	47.4	4.7	942	8	US-08-781-986A-531	Sequence 531, Appl1
10	47.4	4.7	942	16	US-10-329-624-531	Sequence 531, Appl1
11	47.4	4.7	158001	16	US-10-211-179-11	GENERAL INFORMATION
12	47	4.7	549	10	US-09-991-936-879	Sequence 879, Appl1

13	47	4.7	11052	16	US-10-257-166-68	Sequence 68, Appl
14	46.4	2.60	16	US-10-424-599-100851	Sequence 100851,	
15	46.2	4.6	34722	17	US-10-322-281-700	Sequence 700, Appl
16	45.6	4.6	18283	17	US-10-321-613-326	Sequence 326, Appl
17	45.4	4.5	235070	13	US-10-087-192-1990	Sequence 1990, Appl
18	45.2	4.5	4985	14	US-10-094-240-10	Sequence 10, Appl
19	45.2	4.5	4985	15	US-10-056-405-10	Sequence 10, Appl
20	45.2	4.5	13377	15	US-10-311-455-1436	Sequence 1436, Appl
21	45.2	4.5	13377	16	US-10-321-714A-198	Sequence 198, Appl
22	45	4.5	6103	15	US-10-311-455-1664	Sequence 1664, Appl
23	44.8	4.4	1046	13	US-10-027-632-10421	Sequence 10421, A
24	44.8	4.4	1046	15	US-10-027-632-10421	Sequence 10421, A
25	44.8	4.4	3399	18	US-10-793-639-600	Sequence 600, Appl
26	44.6	4.4	973	16	US-10-424-599-58879	Sequence 58879, A
27	44.6	4.4	1488	9	US-09-815-242-4589	Sequence 4589, Ap
28	44.6	4.4	1491	9	US-09-815-242-8559	Sequence 8559, Ap
29	44.6	4.4	1491	6	US-10-282-122A-8174	Sequence 8174, Ap
30	44.6	4.4	13508	16	US-08-781-986A-120	Sequence 120, Appl
31	44.6	4.4	13508	16	US-10-329-624-120	Sequence 120, Appl
32	44.6	4.4	3673778	15	US-10-312-841-1	Sequence 1, Appl
33	44	4.4	6709	16	US-10-424-599-102915	Sequence 102915,
34	44	4.4	18683	15	US-10-311-455-285	Sequence 285, Appl
35	44	4.4	18683	15	US-10-240-452-33	Sequence 33, Appl
36	44	4.4	3673778	15	US-10-312-841-2	Sequence 2, Appl
37	43.8	4.3	567775	17	US-10-741-601-5659	Sequence 5659, Appl
38	43.8	4.3	3673778	15	US-10-312-841-1	Sequence 1, Appl
39	43.6	4.3	6418	16	US-10-311-455-296	Sequence 296, Appl
40	43.6	4.3	6418	15	US-10-321-613-30	Sequence 30, Appl
41	43.4	4.3	3885	18	US-10-377-636-3	Sequence 3, Appl
42	43.4	4.3	4985	14	US-10-094-240-10	Sequence 10, Appl
43	43.4	4.3	4985	15	US-10-056-405-10	Sequence 10, Appl
44	43.4	4.3	11805	15	US-10-311-455-1721	Sequence 1721, Appl
45	43.2	4.3	564	17	US-10-437-963-7115	Sequence 7115, Appl

## ALIGNMENTS

## RESULT 1

```

Sequence 2 Application US/10617962
Publication No. US20040055036A1
GENERAL INFORMATION:
APPLICANT: Commonwealth Scientific and Industrial Research Organisation
APPLICANT: EAST, Peter David
TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus a
FILE REFERENCE: 050179-0076
CURRENT APPLICATION NUMBER: US/10/617,962
CURRENT FILING DATE: 2003-07-14
PRIOR APPLICATION NUMBER: US/09/463,048A
PRIOR FILING DATE: 2002-12-13
PRIOR APPLICATION NUMBER: PCT/AU98/00562
PRIOR FILING DATE: 1998-07-17
PRIOR APPLICATION NUMBER: PO 8088
PRIOR FILING DATE: 1997-07-17
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patentin version 3.1
SEQ ID NO 2
LENGTH: 1008
TYPE: DNA
ORGANISM: Photorhabdus luminescens
US-10-617-962-2

```

	Query Match	Best Local Similarity	100.0%;	Score 1008;	DB 16;	Length 1008;
	Matches 1008;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	ATGGTTATACATTAACACCCGATGATGATGATATCCACCCGTTGAAAACCAATA	60			
Db	1	ATGGTTATACATTAACACCCGATGATGATGATATCCACCCGTTGAAAACCAATA	60			
Qy	61	GCACGAGATATAGTACGTATCTAAACCTTAAGCAAAACAGTAGGGCTATACAGCATCA	120			

Db 61 GCAGAGATATAGTACGTATCTAACTTTAAGCAAAACAGATGAGGGTCAATACGATCA 120

QY 121 TATGGAATGTGAATATCGAGCTPAAGAAAATAATATTAGCTTACGCTTTGGCTGAAGTGT 180

Db 121 TATGGAATGTGAATATCGAGCTPAAGAAAATAATATTAGCTTACGCTTTGGCTGAAGTGT 180

QY 181 ATTCAATATGATATCTAAACCTTCCTGATGACTATTAAGAAATTAAGAACTGCGTAGAGA 240

Db 181 ATTCAATATGATATCTAAACCTTCCTGATGACTATTAAGAAATTAAGAACTGCGTAGAGA 240

QY 241 ATTTATCAAGATATATGTCTAATCTTTCATCTGCACTATTAGGTAAGAAATGTGATCAA 300

Db 241 ATTTATCAAGATATATGTCTAATCTTTCATCTGCACTATTAGGTAAGAAATGTGATCAA 300

QY 301 ATTTCTAAAGATATGCAAAATGGTTTTTATTAAGAATGAATCTGAATTTTGAAGTCAATAT 360

Db 301 ATTTCTAAAGATATGCAAAATGGTTTTTATTAAGAATGAATCTGAATTTTGAAGTCAATAT 360

QY 361 CCTCAAAAACATTTGGAATGTTCTGAGCTTGAAAATAACATGAGTCTTATTTCAGAT 420

Db 361 CCTCAAAAACATTTGGAATGTTCTGAGCTTGAAAATAACATGAGTCTTATTTCAGAT 420

QY 421 GACGATAAATATATAGCACTAATTTTTTCTCTGACGAGAAATTTCCACTGAGAGAAAT 480

Db 421 GACGATAAATATATAGCACTAATTTTTTCTCTGACGAGAAATTTCCACTGAGAGAAAT 480

QY 481 CAAACATCAAAATGCGCAAGATTTTTTAAATTAATGATTTCTTATTACCTTATCTGCT 540

Db 481 CAAACATCAAAATGCGCAAGATTTTTTAAATTAATGATTTCTTATTACCTTATCTGCT 540

QY 541 GTAACTTCACTGCGAAGAGATTTTTTCAAAAACTTTTCAATGATTTAGAGGCTTAA 600

Db 541 GTAACTTCACTGCGAAGAGATTTTTTCAAAAACTTTTCAATGATTTAGAGGCTTAA 600

QY 601 TCATTAGGAATTAATTTAGAGAAAAAACTTTCTAAACCTTTCTTTCGACCAACCGCAG 660

Db 601 TCATTAGGAATTAATTTAGAGAAAAAACTTTCTAAACCTTTCTTTCGACCAACCGCAG 660

QY 661 AGATTACCTGATGCGAGATAGTTATTTGGCTGGAACCAAGAAACGCTAAATGAGAGA 720

Db 661 AGATTACCTGATGCGAGATAGTTATTTGGCTGGAACCAAGAAACGCTAAATGAGAGA 720

QY 721 GTGAGTTTAAAGAACTTAAATAAATACTAGAAATGGAATTTCTAATATGGAAGG 780

Db 721 GTGAGTTTAAAGAACTTAAATAAATACTAGAAATGGAATTTCTAATATGGAAGG 780

QY 781 GCTGCAAAAACAAAGTATAGTTTCAATTTTAAAGAGGTACAAAGGGTAAAGGCTCCACAG 840

Db 781 GCTGCAAAAACAAAGTATAGTTTCAATTTTAAAGAGGTACAAAGGGTAAAGGCTCCACAG 840

QY 841 ACAGCAGCGAAAGATATGTTGTAACGACGAGTGCAGTAACCTGGAATAATTCGGAATAT 900

Db 841 ACAGCAGCGAAAGATATGTTGTAACGACGAGTGCAGTAACCTGGAATAATTCGGAATAT 900

QY 901 TTATATAGTGAAGCTAAGCCAAATAAGACAGGGTAACTTTACTCAAAATGATATGAC 960

Db 901 TTATATAGTGAAGCTAAGCCAAATAAGACAGGGTAACTTTACTCAAAATGATATGAC 960

QY 961 AATAACAATGACGTTCAATGTTGGAACCTCAATTAATAAAATATATGA 1008

Db 961 AATAACAATGACGTTCAATGTTGGAACCTCAATTAATAAAATATATGA 1008

RESULT 2

US-10-617-962-1

; Sequence 1, Application US/10617962

; Publication No. US20040055036A1

; GENERAL INFORMATION:

; APPLICANT: Commonwealth Scientific and Industrial Research Organisation

; APPLICANT: EAST, Peter David

; TITLE OF INVENTION: Toxin Genes from the Bacteria *Xenorhabdus nematophilus* and *Photobacterium*

; TITLE OF INVENTION: *luciferase*

Query Match	Best Local Similarity	79.4%	Score 663.8	DB 16	Length 1107
Matches 835	Conservative 0	Mismatches 162	Indels 54	Gaps 2	
<pre> FILE REFERENCE: 050179-00076 CURRENT APPLICATION NUMBER: US/10/617,962 CURRENT FILING DATE: 2003-07-14 PRIOR APPLICATION NUMBER: US/09/463,048A PRIOR FILING DATE: 2002-12-13 PRIOR APPLICATION NUMBER: PCT/AU98/00562 PRIOR FILING DATE: 1998-07-17 PRIOR APPLICATION NUMBER: PO 8088 PRIOR FILING DATE: 1997-07-17 NUMBER OF SEQ ID NOS: 20 SOFTWARE: PatentIn version 3.1 SEQ ID NO 1 LENGTH: 1107 TYPE: DNA ORGANISM: Xenorhabdus nematophilus US-10-617-962-1 </pre>					
QY	12	ATTAAACCTGATGATAGAGTGGATATCCACCCGTTGAAAAACAATATGACGAGATAT	71		
DB	57	AGTAACGCGCTGATGATTAAGGTGATATCAACCCGTTAAAAAGCAAAATAGCCGGAGATAT	116		
QY	72	AGTAGCTATCTAAACCTTAAGCAAAACGATGAGGGTGATACAGCATCATATGAAATTA	131		
DB	117	AATACGTATCTAGAAATTCAGACAAACAATGAAAGTCAATACAGATTTGATGAAATTC	176		
QY	132	ATATGAGCTAAGAAAAATATATATAGCTTACGCTTGGCTGTAAGTGTATTCATATGT	191		
DB	177	ATATGAGCTAAGAAAGATATATATAGCATATATGCTTACCGGTAAGTGTATTCATATGT	236		
QY	192	ATCTAACTTCTGATGATCTATTAATAAATAAGAGACTGCTGAGAAATTTATCAAGA	251		
DB	237	CTCTCAATTTCCAGAAAGCTATATTAATAAATAAGAGATACACAGGTAGATTTATCAAGA	296		
QY	252	ATTATGCTATCTATCTTTCATCTGCACTATTAGTGAAATGGTGTATCAAAATTTCTAAAGA	311		
DB	297	ATATGCTATCTATCTTTCATCTGCACTATTAGTGAAATGGTGTATCAAAATTTCTAAAGA	356		
QY	312	TATGGCAATGTTTTTATTAAGAATGAATCTGATTTGAAGTCAATATCTCAAAACAT	371		
DB	357	TATGGCAATGATTTTATCCAGAAAGCACTGAGTTTGAAGTCAACGCTTAAAAATATAC	416		
QY	372	TTGGAATGTTCTGAGCTTGAATAAACAATGAGTCTTATTAAGATGACGATTAAT	431		
DB	417	CTGGATATTTCTGATCTTGAATAAACAATGAGTCTTATTAAGATGACGATTAAT	476		
QY	432	ATTAGCACTATATTTTTCTCTGTACAGGAATTCACCTGAGGAAAAATCAACAATCAAA	491		
DB	477	ATTAGCACTATATTTCTTGTCTCAAGAACTTCCATGAGGCAATCAACAATCAAA	536		
QY	492	TGCGCAAGATTTTTTAATTAATTAATGATTTCTTATTAACCTTATCTGCTAATCTTCACT	551		
DB	537	TGAGCAAAATTTTTTAAGATTAATGATTTTATTTACTTATCTTATCTGCTAATCACTACT	596		
QY	552	GGGAAGAGATTTTTTCAAAAAACCTTACAAATGAAATTAAGGCTAATCACTTAGAGA	611		
DB	597	GGGAAGAGATTTTTTCAAAAAATTTTACAAATGCTAGAAATCTAATCACTTAGAGA	656		
QY	612	TTATATGAGGAAAAAACTTTCTAAACCTTTCTTTCGACCAACGACAGAGATTAACCTGA	671		
DB	657	TTATATGAGGAAAAAACTTTCTAAACCTTTCTTTCGACCAACGACAGAGATTAACCTGA	716		
QY	672	TGCGAATATAGTTATTTGGCTGACCAACGAGGCGCTTAA-----714			
DB	717	TGCGAAGACAGGCTTACCTGCGCGGTCCAAACAAAGCGCTTAAATGCAACAGCTCTTC	776		
QY	715	-----TGAGAGGAGTTTTTAAAGACTTAA740			
DB	777	TACAGCAACAGCTCTACAGAGCTTATCTAATTTAGAGAGTTAGTTGCAAAAACCTTG836			



QY 741 AATAACAAATCTAGATGATTTTCTAATATGGAAGGGCTGCAGAAAACAAAGTATAG 800  
 DB 837 AGATTAACCATTCAGAAATACATTTATGAAAAATGATATCTGCAGAAAACAAATATAG 896  
 QY 801 TTCAATTTTAAAGAGGTACAAAAGGGTACGCTTCACAGACAGCAGGAAAGTATTGG 860  
 DB 897 TTCAATTTTAAAGAGGTACAAAAGGGTATGATCAGTGACAGCAGCAGCAATATTGG 956  
 QY 861 TACAGCCAGTGGCAGTACCTCGGAAAAATTCGCCAATA---ATTATATAGTGGAGCT 917  
 DB 957 TACAAAAAGCGCGAGTAACTTCGAAAACTGCAGGTGAGATTTATATATAGTAAAGCT 1016  
 QY 918 AAGCCAAAAGACAGGGGTACCTTACTCAAAATGATCTGCAATATACATGACGGTTCA 977  
 DB 1017 AAGCCAAAGACAGGGGTACCTTACTCAATATATATCTGACCAAAATATATGAGATCCA 1076  
 QY 978 TAGTGTGGAACCTCATTTATAAAAATATATGA 1008  
 DB 1077 AAGTGTGGAACCTCATTTACCAAAATATATATGA 1107

## RESULT 3

US-10-617-962-5  
 ; Sequence 5, Application US/10617962  
 ; Publication No. US20040055036A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
 ; APPLICANT: East, Peter David  
 ; TITLE OF INVENTION: Toxin Genes from the Bacteria *Xenorhabdus nematophilus* and *Photobacterium*  
 ; TITLE OF INVENTION: *luminescens*  
 ; FILE REFERENCE: 050179-0076  
 ; CURRENT APPLICATION NUMBER: US/10/617,962  
 ; CURRENT FILING DATE: 2003-07-14  
 ; PRIOR APPLICATION NUMBER: US/09/463,048A  
 ; PRIOR FILING DATE: 2002-12-13  
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00562  
 ; PRIOR FILING DATE: 1998-07-17  
 ; PRIOR APPLICATION NUMBER: PO 8088  
 ; PRIOR FILING DATE: 1997-07-17  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: Patent version 3.1  
 ; SEQ ID NO 5  
 ; LENGTH: 1205  
 ; TYPE: DNA  
 ; ORGANISM: *Xenorhabdus nematophilus*  
 ; US-10-617-962-5

Query Match 65.9%; Score 663.8; DB 16; Length 1205;  
 Best Local Similarity 79.4%; Pred. No. 2,2e-127;  
 Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;

QY 12 ATTAACACCTGATGATAGAGTATATCCACCGTTGAAAACAAATATGACGAGATAT 71  
 DB 73 AGTAACGCGCTGATGATTAAGGATGATATCAACCGTTGAAAACAAATATGCGGAGATAT 132  
 QY 72 AGTATGATATCTAACTTTAAGCAACAGATGAGGTCATACAGCATCTATATGAAATGA 131  
 DB 133 AATAGGTGATCTAGAAATTCACAAACAAATGAAAGTCTATACAGGATTTGATGAAATGC 192  
 QY 132 ATATGAGCTAGAAAATATATATATAGCTTACGCTTTGGCTGATAGGATTCATATAT 191  
 DB 193 ATATGAGCTAGAAAATATATATATAGCTTACGCTTTGGCTGATAGGATTCATATAT 252  
 QY 192 ATCTAACTCTCTGATGATCTATATATAGAAATTAAGAGACTGCTGAGAGATTTATCAGA 251  
 DB 253 CTCTCAACTTCAGAAAGCTATATATTAATTAAGATTAACAGAGTAGAATTTATCAGAA 312  
 QY 252 ATATATGCTATATCTTATCTGACCTATATAGTGAAGATTTGAGATCAATTTCTAAAGA 311  
 DB 313 ATATATGCTATATCTTATCTGACCTATATAGTGAAGATTTGAGATCAATTTCTAAAGA 372  
 QY 312 TATGCAAAATGCTTTTATATAGAAATGAACTGATTTTGAAGGTCATATCTCAAAACAT 371

DB 373 TATGCAAAATGATTTTATCCAGAAACGACTGGAGTTGGAGTCAACCTTTAAAAATAC 432  
 QY 372 TTGAATGTTCTCGAGCTGAAAATTAACATGAGTGTCTATTCAGATGACGATTAAT 431  
 DB 433 CTGGATATTTCTGATCTTGGAATTAATCTATTTGGAAGATTTATTCAGATGAAATTAAT 492  
 QY 432 ATTAGCACTATATTTTCTCTGACAGAAATTCATCTGAGAGAAATCAACATCAAA 491  
 DB 493 ATTAGCACTATATTTCTTGTCTGACAGAACTTCATATGAGGAAATCAACATCAAA 552  
 QY 492 TGGCGAAGATTTTAAATTAATGATTTCTTATTTACCTTATCTGCTGTAACCTCT 551  
 DB 553 TGACGCAAAATTTTAAAGTAAATGATTTTAACTTATCTTATCTGCTGTAACCTCT 612  
 QY 552 GGGAAAGAGATTTTTCAAAACCTTTACAAAGATTTAGAGGTAATATCATAGAGAA 611  
 DB 613 GGGAAAAAGATTTTTCAAAACCTTTACAAAGATTTAGAGGTAATATCATTTAGAGAA 672  
 QY 612 TTATATTTGAGAGAAAAAACTTTCTTAAACCTTTCTTTCGACCAAGCAGAGATTA 671  
 DB 673 TTATATTTGAGAGAAAAAACTTTCTTAAACCTTTCTTTCGACCAAGCAGAGATTA 732  
 QY 672 TGGCAAAATGATTTTGTGCTGACCAAGCAGAGATTTAA----- 714  
 DB 733 TGGCAAAAGAGCTTATTTGCGCGGTCCAAAGCGCTTAAATTTGCCAACACGCTTTC 792  
 QY 715 -----TGGAGAGTGAAGTTTAAAGACTTAA 740  
 DB 793 TACAGCAACAGCTTACAGAGCTTCACTTAAATTTGAGAGTATTTTGCAAAACCTTAG 852  
 QY 741 AATAACAAATCTAGATGATTTTCTAATATGGAAGGGCTGCAGAAAACAAAGTATAG 800  
 DB 853 AGATTAACCATTCAGAAATACATTTATGAAAAATGATATCTGCAGAAAACAAATATAG 912  
 QY 801 TTCAATTTTAAAGAGGTACAAAAGGGTACGCTTCACAGACAGCAGGAAAGTATTGG 860  
 DB 913 TTCAATTTTAAAGAGGTACAAAAGGGTATGATCAGTGACAGCAGCAGCAATATTGG 972  
 QY 861 TACAGCCAGTGGCAGTACCTCGGAAAAATTCGCCAATA---ATTATATAGTGGAGCT 917  
 DB 973 TACAAAAAGCGCGAGTAACTTCGAAAACCTGCAGAGTATGATTTATATATAGTAACT 1032  
 QY 918 AAGCCAAAAGACAGGGGTACCTTACTCAAAATGATCTGCAATATACATGACGGTTCA 977  
 DB 1033 AAGCCAAAGACAGGGGTACCTTACTCAATATATATCTGACCAAAATATATGAGATCCA 1092  
 QY 978 TAGTGTGGAACCTCATTTATAAAAATATATGA 1008  
 DB 1093 AAGTGTGGAACCTCATTTACCAAAATATATATGA 1123

## RESULT 4

US-10-235-192A-27  
 ; Sequence 27, Application US/10235192A  
 ; Publication No. US20040043389A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: McCarthy, Jeanette  
 ; TITLE OF INVENTION: Methods and Compositions for Identifying  
 ; TITLE OF INVENTION: Risk Factors for Abnormal Lipid Levels and the Diseases  
 ; TITLE OF INVENTION: and Disorders Associated Therewith  
 ; FILE REFERENCE: MMI-011  
 ; CURRENT APPLICATION NUMBER: US/10/235,192A  
 ; CURRENT FILING DATE: 2002-09-04  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 27  
 ; LENGTH: 188971  
 ; TYPE: DNA  
 ; ORGANISM: *Homo sapiens*  
 ; NAME/KEY: misc feature  
 ; LOCATION: 145329-145428, 187896-187995, 195894-195993  
 ; OTHER INFORMATION: N = any nucleotide

US-10-235-192A-27

Query Match 5.0%; Score 50.6; DB 16; Length 188971;  
Best Local Similarity 42.6%; Pred. No. 9.3;  
Matches 263; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

QY 12 ATTAACCTGATGATGAGAGTATTCACCCCTTGAAAAGCAATAGCAGAGATTT 71  
DB 103629 ATCTATATATGAAATATATATGAAATATCTATACAGAAATCTATATCTATAT 103688  
QY 72 AGTAGCTATCTAAACCTTAAGCAACAGATGAGGCTCATACAGCATCATGGAATTA 131  
DB 103689 ATGAATATCTATATGAAATCTATATATGAAATCTATATATCTATATGAAATCT 103748  
QY 132 ATATGAGCTAAGAAATATATATATGAAATATGAAATATGAAATATCTATATGAAAT 191  
DB 103749 ATATGAAATATCATGAAATATATGAAATATGAAATATGAAATATCTATATGAAAT 103808  
QY 192 ATCTAACTCTCTGATGACTATATATGAAATTAAGAGACTGCTGAGAGAAATTTATCA 251  
DB 103809 ATGAATATCTATGAAATATCTATATGAAATATGAAATATCTATATGAAATATATG 103868  
QY 252 ATATATGCTAACTCTTCTGCTGCACTATTTAGTGAAATGGAATCAATTTCTAAAGA 311  
DB 103869 AATTTATATGAAATTTATATATGAAATATATATGAAATATATATGAAATTTATATGA 103928  
QY 312 TATGCAAAATGCTTTTATATAGAAATGAACTGAAATTTGAAAGTCAATATCTCAAAA 371  
DB 103929 TATATATATGAAATATATATATGAAATATATATGAAATATATATATGAAATAT 103988  
QY 372 TTGGAATGCTCTGAGCTTGAAATTAACCATTTGAGTCTTATGAGATGAGATTAAT 431  
DB 103989 TATATATGAAATATATATATGAAATATATATGAAATATATATGAAATATATAT 104048  
QY 432 ATTACACATATATTTTCTCTGTCAGAGAAATTCACATGAGAAATCAACAATCANA 491  
DB 104049 TATATGAAATATATATGAAATATATATGAAATATATATGAAATATATATATAT 104108  
QY 492 TGCCCAAGATTTTAAATTAATGATTTCTTAATTTACCTTATCTGCTGTAACCTG 551  
DB 104109 TGAATTAATATGAAATATATATATGAAATATATATATGAAATATATATATATGA 104168  
QY 552 GGAAGAGAGATTTTTCAAAACTTTACAAATGATTAAGCGCAATCTTAGAGAA 611  
DB 104169 ATTAATATGAAATATATATATGAAATATATATGAAATATATATATATATGAAAT 104228  
QY 612 TTATATGAGAGAAAA 628  
DB 104229 ATATATGAAATATATATAT 104245

RESULT 5  
US-10-240-453-95

; Sequence 95, Application US/10240453  
; Publication No. US20030148326A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
; TITLE OF INVENTION: Transcription  
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated  
; FILE REFERENCE: 5013.1009  
; CURRENT APPLICATION NUMBER: US/10/240,453  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/03973  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 350  
; SEQ ID NO 95  
; LENGTH: 11394

; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-453-95

Query Match 4.9%; Score 49.8; DB 15; Length 11394;  
Best Local Similarity 50.6%; Pred. No. 4.8;  
Matches 120; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 124 GGAATGGAATTCGAGCTAGAAATATATATAGCTTGGCTTGGCTGTAAGTGATTT 183  
DB 318 GGAATGGAATTCGAGCTAGAAATATATATAGCTTGGCTTGGCTGTAAGTGATTT 377  
QY 184 CATATGATCTAAACCTTCTGATGACTATATATAGAAATTAAGAGACTGCTGAGAAAT 243  
DB 378 AT 437  
QY 244 TATCAAGATATATGCTTCAATCTTCACTGCACTATAGGTGAAATGCTGATCAAT 303  
DB 438 AAGGCTTGATTTTAT 497  
QY 304 TCTAAGATATGGAATGCTTTTATATAGAAATGAACTGAAATTTGAAAGTCAATAT 360  
DB 498 TGAATTTTATATGAAATGATTTGATTTAGCTTTTGGCTTTTAAATAT 554

RESULT 6

US-10-349-680-73  
; Sequence 73, Application US/10349680  
; Publication No. US20030176654A1

; GENERAL INFORMATION:  
; APPLICANT: Casselli, Gail  
; APPLICANT: Chen, Ellison  
; APPLICANT: Glass, Jennifer  
; APPLICANT: Glass, John  
; APPLICANT: Heiner, Cheryl  
; APPLICANT: Lefkowitz, Elliott  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA  
; FILE REFERENCE: UAB-13403/22  
; CURRENT APPLICATION NUMBER: US/10/349,680  
; CURRENT FILING DATE: 2003-01-23  
; PRIOR APPLICATION NUMBER: US 09/601,198  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: PCT/US99/01972  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: US 60/073,189  
; PRIOR FILING DATE: 1998-01-30  
; NUMBER OF SEQ ID NOS: 181  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 73  
; LENGTH: 1866  
; TYPE: DNA  
; ORGANISM: Ureaplasma urealyticum  
US-10-349-680-73

Query Match 4.8%; Score 48.8; DB 15; Length 1866;  
Best Local Similarity 47.3%; Pred. No. 4;  
Matches 249; Conservative 0; Mismatches 267; Indels 10; Gaps 3;

QY 122 ATGAATGGAATATGAGCTAGAAATATATATAGCTTACGCTTGGCTGTAAGTGAT 181  
DB 1037 ATGAATGGAATATGAGCTAGAAATATATATAGCTTACGCTTGGCTGTAAGTGAT 1096  
QY 182 TTCAATGATCTAACTTCTGATGACTTTATATAGAAATTAAGAGCTGCTGAGAA- 240

Db 1097 TTCAAAATATATGATCATATATTTAGATATTAATTTAATTAATTTCAAGAGCAT 1156  
Qy 241 --ATTATCAAGATATATATGCTATCTTCACTGCACTATAGTGAATGGTGATC 298  
Db 1157 ATATTACGAAAAATTAATTAATTAATTTATGAGCTTTCATATATATGATGTC 1216  
Qy 239 AAATTTCTAAGATATGCAAAATGGTTTTTATAGAAATGAATGATTTTGAAGTCAAT 358  
Db 1217 AAATTACAAAATATCTAATTTGACAAAATGATAGAAATACCAAAATTAATTAATTAAC 1276  
Qy 339 ATCTCAAAAACCTTTGGAAATGTTCTGAGCTTGAATAATTAACATGAGTGTATTGAG 418  
Db 1277 AATTATATGAAATTTAGTATATTTTAAATTAATTAATTAATTAATTAATTAAT 1336  
Qy 419 ATGAGATTAATTTT-----ATTAGCACTATATTTTCTCTGACAGAAATTCACCTGGA 473  
Db 1337 ATAAACCAAAATTAATGATTTATGAAAAATTTTTCACAAAATGGTATTTTATTTAA 1396  
Qy 474 GGAATATCAACATCAAAATGCGCAAGATTTTAAATTAATTAATTAATTAATTAAT 533  
Db 1397 TGAATTTAAGCA--CTTGAATTAAGTTTAAATTAATTAATTAATTAATTAATTAAT 1454  
Qy 534 ATCTGCTGTAATCTTCACTGGGAGAGGATTTTTCACAAAATCTTTTCAATGATTAAG 593  
Db 1455 AAATCTTATTAATAAGATGTTAGACCAAGTGAATAATTTTAAAGGTTTAAAAA 1514  
Qy 594 GGCTAAATCATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 639  
Db 1515 ACTTAACCATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 1560

## RESULT 7

US-10-301-533-23  
; Sequence 23, Application US/10301533  
; Publication No. US20030203473A1  
; GENERAL INFORMATION:  
; APPLICANT: Godzik, Adam  
; APPLICANT: Reed, John C.  
; TITLE OF INVENTION: Microbial Sumo Protease Homologs  
; FILE REFERENCE: P-LD 5488  
; CURRENT APPLICATION NUMBER: US/10/301,533  
; CURRENT FILING DATE: 2002-11-20  
; PRIOR APPLICATION NUMBER: US 60/331,895  
; PRIOR FILING DATE: 2001-11-20  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: FaecSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 2767  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2) ... (1507)  
US-10-301-533-23

Query Match 4.8%; Score 48.2; DB 15; Length 2767;  
Best Local Similarity 56.7%; Pred. No. 6.1;  
Matches 89; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 243 TTATCAAGATATATGCTAATCTTTCATCTGCACTATAGTGAATATGATCAAT 302  
Db 562 TCATCTAGATTAATAATTCCTGCTTTCTAATTAATTAATTAATTAATTAATTAAT 621  
Qy 303 TTCTAAGATATGCAAAATGGTTTTTATAGAAATGAATGATTTTGAAGTCAATATCC 362  
Db 622 GTGTATGAAATGAAGATCATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 681  
Qy 363 TCAAAACATTTGGAATGCTCTGAGCTTGAATAAA 399  
Db 682 TGATGAGTATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 718

## RESULT 8

US-10-311-455-597  
; Sequence 597, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEBENDROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
; FILE REFERENCE: 5013, 1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 597  
; LENGTH: 16217  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 12269  
; OTHER INFORMATION: n is a or g or c or t  
US-10-311-455-597

## Query Match 4.7%; Score 47.6; DB 15; Length 16217;

Best Local Similarity 43.3%; Pred. No. 16;  
Matches 221; Conservative 0; Mismatches 289; Indels 0; Gaps 0;

Qy 124 GGAATGGAATATCGAGCTAAGAAATATATTAATTAATTAATTAATTAATTAATTAAT 183  
Db 2495 GAAATATGATATTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 2554  
Qy 184 CATATGATCTAATTAATCTCGATGATCTATTAAGATTAAGAGATCTGAGAGATTT 243  
Db 2555 AAGAAAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2614  
Qy 244 TATCAAGATATATGCTAATCTTTCATCTGCACTATAGTGAATATGATCAATTT 303  
Db 2615 TGGGTTGTTTATGTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAAT 2674  
Qy 304 TCTAAGATATGCAAAATGGTTTTTATAGAAATGAATGATTTGAAGTCAATATCT 363  
Db 2675 TATATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTA 2734  
Qy 364 CAAAACATTTGGAATGTTCTGAGCTTGAATAATAACATTAAGATGATCAATGAC 423  
Db 2735 AAGGTAATTAATTAATTTGTTATTAATTAATTAATTAATTAATTAATTAATTAAT 2794  
Qy 424 GATAATTAATTAAGCACTAATTTTCTCTGACAGAAATTCACGAGAGAAATCA 483  
Db 2795 AGTATTTAGTTAGTTTATTAATTAATTAATTTGTTATTAATTAATTAATTAATTAAT 2854  
Qy 484 CAATCAATATGCGCAAGATTTTAAATTAATTAATTTCTTAATTAATTAATTAATTAAT 543  
Db 2855 TTTTATTAATTAAGAAATATTTGTTGTTGTTAGTTTATTAATTAATTAATTAATTAAT 2914  
Qy 544 ACTTCACTGGAGAGAGATTTTCAAAAACCTTTCAATGATTAAGAGCTTAATCA 603  
Db 2915 AGTATGATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2974  
Qy 604 TTAGAGATTAATTAAGAGAAAAAATCTT 633  
Db 2975 AGTGGATTTTATTAAGAGATTAATTAATTT 3004

## RESULT 9

US-08-781-986A-531  
; Sequence 531, Application US/08781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunesh  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 531:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 942 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-781-986A-531

Query Match 4.7%; Score 47.4; DB 8; Length 942;  
Best Local Similarity 47.8%; Pred. No. 6;  
Matches 138; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 66 AGATATAGTACGTAACTTAACCTTAAGCAACAGATGAGGTCAATACATCATATG 125  
DB 313 AGAAAACGTACTTAAGCAATGAGCAATTAAGTAAGCAATTAAGTCACTTCTAT 372  
QY 126 AATTGATATTCAGCTAAGAAAATATATTAGCTTACGCTTGGCTGTAAGGTATTC 185  
DB 373 AGATGAATTAAGAGCTTAAGCTAAGTATCGTATATTCATGATGGTGAACGCTG 432  
QY 186 TAATGATATTAACCTTCGATGACATATATAAGAAATAAGACCTGCGAGAGATT 245  
DB 433 CATAGAGAGAAATTAATATGATGACATATGCAATGAAGATATTTTAAAGAAACGA 492  
QY 246 TCAAGAAATATATGCTTAATCTTCACTGCACTATTAGTGAAATGATCAATTTTC 305  
DB 493 AACGCATCAATGCTGATTTTGTATTTAGTATATGATGATCAATGAACCTATGTT 552  
QY 306 TAAAGATATGCAAAATGTTTTTATTAAGATGAATGAACTGGAATTTGAAGT 354  
DB 553 AAAACAGAGACATACGCTTTTATTAAGAAAGAAATCTATTAAGAT 601

RESULT 10  
US-10-329-624-531  
; Sequence 531, Application US/10329624  
; Publication No. US20040043037A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunesh  
; GIL H. Choi

Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5256  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/329,624  
; FILING DATE: 27-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/956,171  
; FILING DATE: October 20, 1997  
; APPLICATION NUMBER: 60/009,861  
; FILING DATE: January 5, 1996  
; APPLICATION NUMBER: 08/781,986  
; FILING DATE: January 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Hyman  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB248P1D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (240) 314-1224  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 531:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 942 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 531:  
; US-10-329-624-531

Query Match 4.7%; Score 47.4; DB 16; Length 942;  
Best Local Similarity 47.8%; Pred. No. 6;  
Matches 138; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 66 AGATATAGTACGTAACTTAACCTTAAGCAACAGATGAGGTCAATACATCATATG 125  
DB 313 AGAAAACGTACTTAAGCAATGAGCAATTAAGTAAGCAATTAAGTCACTTCTAT 372  
QY 126 AATTGATATTCAGCTAAGAAAATATATTAGCTTACGCTTGGCTGTAAGGTATTC 185  
DB 373 AGATGAATTAAGAGCTTAAGCTAAGTATCGTATATTCATGATGGTGAACGCTG 432  
QY 186 TAATGATATTAACCTTCGATGACATATATAAGAAATAAGACCTGCTGAGAGATT 245  
DB 433 CATAGAGAGAAATTAATATGATGACATATGCAATGAAGATATTTTAAAGAAACGA 492  
QY 246 TCAAGAAATATATGCTTAATCTTCACTGCACTATTAGTGAAATGATCAATTTTC 305  
DB 493 AACGCATCAATGCTGATTTTGTATTTAGTATATGATGATCAATGAACCTATGTT 552  
QY 306 TAAAGATATGCAAAATGTTTTTATTAAGATGAATGAACTGGAATTTGAAGT 354  
DB 553 AAAACAGAGACATACGCTTTTATTAAGAAAGAAATCTATTAAGAT 601

RESULT 11  
US-10-211-179-11  
; GENERAL INFORMATION:  
; APPLICANT: Nicholas M. Dean



```

; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 68
; LENGTH: 11052
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-68

```

```

Query Match          4.7%; Score 47; DB 16; Length 11052;
Best Local Similarity 48.7%; Pred. No. 18;
Matches 128; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```

```

QY 225 AGAGCTGCTGAGAGAAATTTTATCAAGATATGCTATCTTTCATCTGCACATTTAG 284
DB 1196 AGAAGCTATGAGATTTATTTTAAAGTTTGGGTTTATGAAAGATATATTTGG 1255
QY 285 TGAAGATGCTGATCAAAATTTCTAAGATATGCAAAATGCTTTTATAGAAATGAAC 344
DB 1256 AGTAAATGTTTAAAGAGAGAAAGGTGAGAAAGGGGTTTATATATGATGATTA 1315
QY 345 TTTGAAGCTCATATCTCCAAATCTTGGATGCTTCTGAGCTTGAATAAACAT 404
DB 1316 TTTAATGATTTTATTTTAAAGGGGTTGGGTGAGGATGATTAATAAAGTT 1375
QY 405 GAGTCTTATTCAGATGACGATTAATTTAGCACTAATTTTCTCTGACAGGAAT 464
DB 1376 AAAAATTTAAATTTTGTATATTTTATGCTTATTTTGTGTTTATAGATTA 1435
QY 465 TCCACTGAGAGAAATCAACAT 487
DB 1436 TTGAGTTTGAATAATTATAGT 1458

```

```

RESULT 14
US-10-424-599-100851
; Sequence 100851, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 100851
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_62081C.1
US-10-424-599-100851

```

```

Query Match          4.6%; Score 46.4; DB 16; Length 260;
Best Local Similarity 52.6%; Pred. No. 6;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```

```

QY 249 AGAATATATGCTATCTTTCATCTGCATATTTAGTGAAATGCTGATCAATTTCTAA 308
DB 1 ATAACTGTTAATCTTTTAAACATTTTAAAGGCAAAAGGACTGAAGTTTCGAA 60
QY 309 AGATATGCAATGCTTTTATAGAAATGACTGATTTTGAAGTCAATATCTCAAA 368
DB 61 AAAAGAGAAATAGTTTCAAGATTTCTTCACTGAAATTTTAAATCAAA 120
QY 369 CATTGGAATGTTCTGAGCTTGAATAAATCAATGATGCTTATTCAGATGACATTA 428

```

```

DB 121 CGTTGATTTTTTAAAAAATGACATTAACAAAGTTTGCTATCTTAAGGGGGAAA 180
QY 429 ATTATGACACT 440
DB 181 AACTTACTACT 192

```

```

RESULT 15
US-10-322-281-700
; Sequence 700, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700
; LENGTH: 34722
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(34722)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-700

```

```

Query Match          4.6%; Score 46.2; DB 17; Length 34722;
Best Local Similarity 45.0%; Pred. No. 40;
Matches 174; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

```

```

QY 50 AAAAGCAATGCGAGATATAGCTATATCTTAATCTTAAGCAACAGATGAGGTC 109
DB 2706 AAAATTTATTAATAAGAGAAACAAATTAATATATATATATATATATATATAT 2765
QY 110 ATACGATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 169
DB 2766 ATAAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTA 2825
QY 170 CTGTAAGTGATTCATATGATCTATTAACCTTCTGATGACTATTAAGATTAAGA 229
DB 2826 ATGATATATATTTAATAATTAATAATATATATATTAATAATTAATAATTAATA 2885
QY 230 CTGCTGAGAGATTTATGAGAAATATATGCTTAATCTTCACTGACCTATTTAG 289
DB 2886 TATGTTTATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 2945
QY 290 ATGCTGATCAAAATTTCTAAAGATATGCAAAATGTTTATTAAGATGATGATTT 349
DB 2946 TTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTAT 3005
QY 350 AAGTCAATATCTGCAAAACATTTGGAATGTTCTGACCTGCAAAATTAACCAT 409
DB 3006 AATATATATATATATTAATAATTAATAATTAATAATTAATAATTAATAATTA 3065
QY 410 CTTATTCAGATGACATTAATTTATG 436
DB 3066 ATTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3092

```

Search completed: November 21, 2004, 08:42:21  
 Job time : 589 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 07:28:02 ; Search time 116 Seconds  
(without alignments)  
6176.510 Million cell updates/sec

Title: US-10-617-962-2  
Perfect score: 1008  
Sequence: 1 atggttatacaatracacc.....ctcatataaataatgta 1008

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	100.0	1008	US-09-463-048A-2	Sequence 2, Appli
2	663.8	65.9	1107	US-09-463-048A-1	Sequence 1, Appli
3	663.8	65.9	1205	US-09-463-048A-5	Sequence 5, Appli
4	628.2	62.3	1272	US-08-569-168-6	Sequence 6, Appli
5	521.4	51.7	834	US-08-569-168-3	Sequence 3, Appli
6	521.4	51.7	837	US-08-569-168-1	Sequence 1, Appli
7	62.6	6.2	1141	US-08-806-708B-22	Sequence 22, Appli
8	59.8	5.9	7218	US-08-232-463-14	Sequence 23, Appli
9	51	5.1	1055	US-09-806-708B-23	Sequence 23, Appli
10	50.6	5.0	1141	US-09-806-708B-22	Sequence 22, Appli
11	48.8	4.8	1866	US-09-601-198-153	Sequence 153, App
12	47.8	4.7	729	US-09-134-001C-1161	Sequence 1161, Ap
13	47.6	4.7	1055	US-08-806-708B-23	Sequence 23, Appli
14	47.4	4.7	942	US-08-956-171E-531	Sequence 531, App
15	47.4	4.7	942	US-08-781-986A-531	Sequence 531, App
16	47.4	4.7	41708	US-09-470-512A-3	Sequence 3, Appli
17	47.4	4.7	41708	US-09-676-519-18	Sequence 18, Appli
18	47	4.7	5340	US-09-627-122-21	Sequence 21, Appli
19	44.8	4.4	3399	US-09-614-221A-600	Sequence 600, App
20	44.6	4.4	13508	US-08-956-171E-120	Sequence 120, App
21	44.6	4.4	13508	US-08-781-986A-120	Sequence 120, App
22	43.8	4.3	832	US-09-621-976-2813	Sequence 2813, App
23	43.8	4.3	665	US-08-883-795A-36	Sequence 36, Appli
24	43.2	4.3	1470	US-09-830-217-1	Sequence 1, Appli
25	42.8	4.2	7425	US-09-453-702B-212	Sequence 212, App
26	42.6	4.2	5340	US-09-627-122-21	Sequence 21, Appli
27	42.6	4.2	640681	US-09-790-988-1	Sequence 1, Appli

28	42.4	4.2	627	US-09-248-796A-14060	Sequence 14060, A
29	42.2	4.2	5136	US-08-956-171E-332	Sequence 332, App
30	42.2	4.2	5136	US-08-781-986A-332	Sequence 332, App
31	41.8	4.1	2557	US-09-710-279-3829	Sequence 3829, Ap
32	41.8	4.1	3146	US-09-710-279-4281	Sequence 4281, Ap
33	41.8	4.1	3646	US-09-710-279-4070	Sequence 4070, Ap
34	41.4	4.1	978	US-09-248-796A-5337	Sequence 5337, Ap
35	41.4	4.1	2424	US-08-956-171E-392	Sequence 392, App
36	41.4	4.1	2424	US-08-781-986A-392	Sequence 392, App
37	41.2	4.1	2142	US-09-107-532A-905	Sequence 905, App
38	41.2	4.1	8920	US-08-446-855A-1	Sequence 1, Appli
39	41.2	4.1	8920	US-09-150-741-1	Sequence 1, Appli
40	41.2	4.1	14066	US-09-601-198-56	Sequence 56, Appli
41	41	4.1	1578	US-09-248-796A-5803	Sequence 5803, Ap
42	41	4.1	43095	US-09-676-519-17	Sequence 17, Appli
43	40.8	4.0	832	US-09-621-976-2813	Sequence 2813, App
44	40.8	4.0	1830121	US-09-557-884-1	Sequence 1, Appli
45	40.8	4.0	1830121	US-09-643-990A-1	Sequence 1, Appli

# ALIGNMENTS

```

RESULT 1
US-09-463-048A-2
; Sequence 2, Application US/09463048A
; Patent No. 6630619
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photoc
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photoc
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/09/463,048A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Photorhabdus luminescens
US-09-463-048A-2

Query Match      100.0%; Score 1008; DB 4; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1.2e-242;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGTTATACATTAACACCTGATGATGAAGTGATATCCACCGTTGAAAGCAATA 60
        1 ATGTTATACATTAACACCTGATGATGAAGTGATATCCACCGTTGAAAGCAATA 60
DB      1 ATGTTATACATTAACACCTGATGATGAAGTGATATCCACCGTTGAAAGCAATA 60
        1 ATGTTATACATTAACACCTGATGATGAAGTGATATCCACCGTTGAAAGCAATA 60
QY      61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGCTGATACGATCA 120
        61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGCTGATACGATCA 120
DB      61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGCTGATACGATCA 120
        61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGCTGATACGATCA 120
QY      121 TATGAATGATATCGAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 180
        121 TATGAATGATATCGAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 180
DB      121 TATGAATGATATCGAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 180
        121 TATGAATGATATCGAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 180
QY      181 ATTCAATGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 240
        181 ATTCAATGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 240
DB      181 ATTCAATGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 240
        181 ATTCAATGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 240
QY      241 ATTATCAAGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 300
        241 ATTATCAAGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 300
DB      241 ATTATCAAGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 300
        241 ATTATCAAGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 300
QY      301 ATTCTTAAGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 360
        301 ATTCTTAAGATATGATTAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGA 360

```

Db 301 ATTCTAAAGATATGCGCAATGCTTTTATTAAGAAATGAACCTGATTTTGAAGTCAATAT 360  
 Qy 361 CCTCAAAACATTGGAAATGCTCTGAGCTTGAATAAACAATGAGTCTTATTCAGAT 420  
 Db 361 CCTCAAAACATTGGAAATGCTCTGAGCTTGAATAAACAATGAGTCTTATTCAGAT 420  
 Qy 421 GACGATAAATTTTGAAGTCTTATTTCTCTGACAGAAATTTCCATGAGGAAAT 480  
 Db 421 GACGATAAATTTTGAAGTCTTATTTCTCTGACAGAAATTTCCATGAGGAAAT 480  
 Qy 481 CAACATCAAAATGCGCAAGATTTTAAATTAATGATTTCTTTTAACTTATCTGCT 540  
 Db 481 CAACATCAAAATGCGCAAGATTTTAAATTAATGATTTCTTTTAACTTATCTGCT 540  
 Qy 541 GTAACTTCACTGGAGAGAGATTTTTCACAAAACCTTTTACAAATGATTTAGAGCTAA 600  
 Db 541 GTAACTTCACTGGAGAGAGATTTTTCACAAAACCTTTTACAAATGATTTAGAGCTAA 600  
 Qy 601 TCATTTAGAAATTTATTTGAGAGAAATTTTCTTAAACCTTTCTTTGACCAACGCGAG 660  
 Db 601 TCATTTAGAAATTTATTTGAGAGAAATTTTCTTAAACCTTTCTTTGACCAACGCGAG 660  
 Qy 661 AGATTACCTGATGAGAGATTTATTTGCTGAGCAAGAGAGCGCTTAAATGAGAGA 720  
 Db 661 AGATTACCTGATGAGAGATTTATTTGCTGAGCAAGAGAGCGCTTAAATGAGAGA 720  
 Qy 721 GTGAGTTTAAAGAACTTAAATAAACAATCTAGAGATTTTCTAATATGAGAGG 780  
 Db 721 GTGAGTTTAAAGAACTTAAATAAACAATCTAGAGATTTTCTAATATGAGAGG 780  
 Qy 781 GGTGAGAAACAAATATTTGCTTATTTAAGAGGTACAAAGGTACCTCTCACAG 840  
 Db 781 GGTGAGAAACAAATATTTGCTTATTTAAGAGGTACAAAGGTACCTCTCACAG 840  
 Qy 841 ACAGAGCGAAAGATTTGCTACAGCAGTGCAGTGAACCTGAAATTTGCGAATAT 900  
 Db 841 ACAGAGCGAAAGATTTGCTACAGCAGTGCAGTGAACCTGAAATTTGCGAATAT 900  
 Qy 901 TTATATGATGAGCTTAAGCCAAAAGACAGGGTAACTTTACTCAAAATGATCTGAC 960  
 Db 901 TTATATGATGAGCTTAAGCCAAAAGACAGGGTAACTTTACTCAAAATGATCTGAC 960  
 Qy 961 AATAAAGAGCGTTCATATGCTTGAACCTCAATTTAATAAATATATGA 1008  
 Db 961 AATAAAGAGCGTTCATATGCTTGAACCTCAATTTAATAAATATATGA 1008  
 RESULT 2  
 US-09-463-048A-1  
 ; Sequence 1, Application US/09463048A  
 ; Patent No. 6630619  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
 ; APPLICANT: EAST, Peter David  
 ; TITLE OF INVENTION: Toxin Genes from the Bacteria *Xenorhabdus nematophilus* and *Photobacterium*  
 ; FILE REFERENCE: 050179-0076  
 ; CURRENT APPLICATION NUMBER: US/09/463,048A  
 ; CURRENT FILING DATE: 2002-12-13  
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00562  
 ; PRIOR FILING DATE: 1998-07-17  
 ; PRIOR APPLICATION NUMBER: PO 8088  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 1  
 ; LENGTH: 1107  
 ; TYPE: DNA  
 ; ORGANISM: *Xenorhabdus nematophilus*  
 US-09-463-048A-1  
 Query Match 65.9%; Score 663.8; DB 4; Length 1107;  
 Best Local Similarity 79.4%; Pred. No. 1.3e-156;

	Matches	835; Conserved	0; Mismatches	162; Indels	54; Gaps	2;
Qy 12	ATTAAACACCTGATGATAGATGATATCAACCGCTTGAAGCAATAGAGAGATAT	71				
Db 57	AGTAAACACCTGATGATAGATGATATCAACCGCTTGAAGCAATAGAGAGATAT	116				
Qy 72	AGTAAACACCTGATGATAGATGATATCAACCGCTTGAAGCAATAGAGAGATAT	131				
Db 117	AGTAAACACCTGATGATAGATGATATCAACCGCTTGAAGCAATAGAGAGATAT	176				
Qy 132	ATATGAGCTTGAAGCAATAGATGATATCAACCGCTTGAAGCAATAGAGAGATAT	191				
Db 177	ATATGAGCTTGAAGCAATAGATGATATCAACCGCTTGAAGCAATAGAGAGATAT	236				
Qy 192	ATCTAACTTCTGATGATGATATGATATGATATGATATGATATGATATGATAT	251				
Db 237	CTCTCACTTCTGATGATGATATGATATGATATGATATGATATGATATGATAT	296				
Qy 252	ATATATGCTTAACTTTTATGATGATATGATATGATATGATATGATATGATAT	311				
Db 297	ATATATGCTTAACTTTTATGATGATATGATATGATATGATATGATATGATAT	356				
Qy 312	TATGCAAAATGATTTTATTAAGATGATGATATGATATGATATGATATGATAT	371				
Db 357	TATGCAAAATGATTTTATTAAGATGATGATATGATATGATATGATATGATAT	416				
Qy 372	TTGAAATGCTTCTGATGATGATATGATATGATATGATATGATATGATATGAT	431				
Db 417	CTGGGATATTTCTGATGATGATATGATATGATATGATATGATATGATATGAT	476				
Qy 432	ATTAGCACTATATTTTCTGATGATGATATGATATGATATGATATGATATGAT	491				
Db 477	ATTAGCACTATATTTTCTGATGATGATATGATATGATATGATATGATATGAT	536				
Qy 492	TGCGCAAGATTTTAAATTAATGATATGATATGATATGATATGATATGATAT	551				
Db 537	TGCGCAAGATTTTAAATTAATGATATGATATGATATGATATGATATGATAT	596				
Qy 552	GGAAGAGAGATTTTCAAAAACCTTTCAATGATGATGATGATGATGATGATG	611				
Db 597	GGAAGAGAGATTTTCAAAAACCTTTCAATGATGATGATGATGATGATGATG	656				
Qy 612	TTATATGAG	671				
Db 657	TTATATGAG	716				
Qy 672	TGCGCAAGATTTTAAATTAATGATATGATATGATATGATATGATATGATAT	714				
Db 717	TGCGCAAGATTTTAAATTAATGATATGATATGATATGATATGATATGATAT	776				
Qy 715	-----TGAGAGTGAAGTTTAAAGAACTTAA 740					
Db 777	TACAGCAACACGCTTACAGCAGCTTCAATTAATGAGAGAGTTTGAAGAACTT	836				
Qy 741	AAATTAACAATCTTGAAGATGATTTTCTAATATGAGAGAGCTGCAAAACAAAT	800				
Db 837	AGATTAACCATTCAGAAATTAATTAATGATATGATATGATATGATATGATAT	896				
Qy 801	TTCAATTTAAAGAGTACAAAGAGGATGATGATGATGATGATGATGATGATG	860				
Db 897	TTCAATTTAAAGAGTACAAAGAGGATGATGATGATGATGATGATGATGATG	956				
Qy 861	TACAGCAGTGGCAGTAACTTGAAGAAATTTCCGAGATA--ATTATATAGTGA	917				
Db 957	TACAGCAGTGGCAGTAACTTGAAGAAATTTCCGAGATA--ATTATATAGTGA	1016				
Qy 918	AAAGCAAAAGACAGAGGATGATGATGATGATGATGATGATGATGATGATG	977				
Db 1017	AAAGCAAAAGACAGAGGATGATGATGATGATGATGATGATGATGATGATG	1076				
Qy 978	TAGTGTGAAGCTCATTAATAAATATATGA 1008					
Db 1077	AAAGTGTGAAGCTCATTAATAAATATATGA 1107					





QY 12 ATTAACCTGTATGATGAGATGATATCCACCGTTGAAAAACAATATGACGAGATAT 71  
DB 139 AGTAACGCTGTATGATGATGAGATATCAACCGTTGAAAAACAATATGACGAGATAT 198  
QY 72 AGTAGCTATCTAACTTTAAGCAACAGATGAGGCTCTATACGATCATATGSAATTGA 131  
DB 199 AATAGCTGTACTAGAAATTCAGCAACAAATGAAAGTCAACAGGATGATGAGAAATTC 258  
QY 132 ATATGAGCTAAGAAATATATATATAGCTTACCTTTGGCTGTATGATGATCATATAT 191  
DB 259 AATGAGCTAAGAAATGATATATATAGCATATGCTTTAGCGGTATGATGATCATATAT 318  
QY 192 ATCTAACTCTCTGATGATCTATATATAGAAATTAAGACTGTCTGAGAGAAATTTATCA 251  
DB 319 CTCTCAACTCTCAGAAAGCTATATATTAATAAGATTAACAGAGTAAATTTATCAAGT 378  
QY 252 ATATATGCTATCTTTGATCTGCACTATAGTGAATATGATGATCAAAATTTCTAAAGA 311  
DB 379 ATACATGCTAATCTTTATCTGCACTATGAGTGAATATGATGATCAAAATTTCTAAAGA 438  
QY 312 TATGCAAAATGCTTTTATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 371  
DB 439 TATGCAAAATGATTTTATCCAGAGAACTGAGATTTG-AGGTCAACGCTTTAAATATC 497  
QY 372 TTGAATGTTCTGAGCTTGAATTAACCATT-AGTGTCTTATTCAGATGATGATGAT 430  
DB 498 CTGGATATTCCTGATCTTGAAGATTAACATATGGAATATTTATTCAGATGATGAT 557  
QY 431 TATGCACTATATTTTCTGCTGCAAGAAATTCCTGAGAGAAATTCAAATCA 490  
DB 558 TATTAAGCTATATATTTCTGCTGCAAGAAATTCCTGAGAGAAATTCAAATCA 617  
QY 491 ATGCGCAAGATTTTATTAATTAATGATTTCTTATTAATCTTATCTGCTGATCTTAC 550  
DB 618 ATGCAAGCAATTTTATTAAGATGATTTTATCTTATCTTATCTGCTGATCTTAC 677  
QY 551 TGGGAAGAGATTTTTCAAAACTTTTCAATGATTAAGAGGCTAAATCATTAAGAGA 610  
DB 678 TGGGAAGAGATTTTTCAAAACTTTTCAATGATTAAGAGGCTAAATCATTAAGAGA 737  
QY 611 ATTATATGAGAGAAAACTTTCTAAACCTTTCTTGAACACCGGAGATTAACCTG 670  
DB 738 ATTATATGAGAGAAAACTTTCTAAACCTTTCTTGAACACCGGAGATTAACCTG 797  
QY 671 ATGCAAGATAGTATTTGCTGACCAACAGAGGCTTAA-----714  
DB 798 ATGCAAGACAGGCTATCTTGCGCGTCCAAAGCGCTTAAATTTGCCAAGACGCTT 857  
QY 715 -----TGAGAGTGAATTT-TAAAGACTT 738  
DB 858 CTACAGCAACAGCTCTACAGCAGCTTCATCTAATTTGAGAGTGAATTTGCCAAGACTT 917  
QY 739 AAAATTAACAATCTACAGAAATGATTTTCTAATATGAGAGGCTGCCAAGCAAAAGTAT 798  
DB 918 AGAGATTAACCCATCCAGAAATACATTTATGAAATGATGATGCTGCAAAAGCAAAATAT 977  
QY 799 AGTCAATTTATTAAGAGGCTACAAAAAGGCTCAACAGACAGCAAGCAAAAGTAT 858  
DB 978 AGTCAATTTATTAAGAGGCTACAAAAAGGCTCAACAGACAGCAAGCAAGTAT 1037  
QY 859 GGTACAGCAGTGGCAGTAACTGAAAAATTTGCCGAATA--ATTATATATGCTGAG 915  
DB 1038 GGTACAAAAAGCGCAGTAACTTGCAAAACTGCAAGGTATGATTTATATGATTAAGA 1097  
QY 916 CTAAACCAAAAAAGAGGCTTAACTTCAAAAATGATATGACAAATCATGACGCTT 975  
DB 1098 CTAAACCAAAAAAGAGGCTTAACTTCTCAATTAATATATGACCAAAATATGAGATC 1157  
QY 976 CATAGTGTGAACATCATATATATATATATGA 1008  
DB 1158 CAAAGTGTGAACATCATATATATATATGA 1190

RESULT 5  
US-08-569-168-3  
Sequence 3, Application US/08569168  
Patent No. 5972687  
GENERAL INFORMATION:  
APPLICANT: Smigielski, Adam J.  
APPLICANT: Ahnurst, Raymond J.  
TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lowe, Price, Leblanc & Becker  
STREET: 99 Canal Center Plaza, Suite 300  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/569,168  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Price, Robert L.  
REGISTRATION NUMBER: 22,685  
REFERENCE/DOCKET NUMBER: 1451-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-684-1111  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 834 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-569-168-3  
Query Match 51.7%; Score 521.4; DB 2; Length 834;  
Best Local Similarity 85.4%; Pred. No. 4.5e-121;  
Matches 604; Conservative 0; Mismatches 101; Indels 2; Gaps 2;  
QY 12 ATTAACCTGTATGATGAGATGATATCCACCGTTGAAAAACAATATGACGAGATAT 71  
DB 57 AGTAACGCTGTATGATGATGAGATATCAACCGTTGAAAAACAATATGACGAGATAT 116  
QY 72 AGTAGCTATCTAACTTTAAGCAACAGATGAGGCTCTATACGATCATATGSAATTGA 131  
DB 117 AATAGCTGTACTAGAAATTCAGCAACAAATGAAAGTCAACAGGATGATGAGAAATTC 176  
QY 132 ATATGAGCTAAGAAATATATATATAGCTTACCTTTGGCTGTATGATGATCATATAT 191  
DB 177 AATAGCTAAGAAATGATATATATATGATGATGATGATGATGATGATGATGATGAT 236  
QY 192 ATCTAACTCTCTGATGATCTATATATAGAAATTAAGACTGTCTGAGAGAAATTTATCA 251  
DB 237 CTCTCAACTCTCAGAAAGCTATATTAATAATGATTAACAGATTAATTTATCAAGT 296  
QY 252 ATATATGCTAATCTTTATCTGCACTATTAAGTGAATATGATGATGATGATGATGAT 311  
DB 297 ATACATGCTAATCTTTATCTGCACTATGAGTGAATGATGATGATGATGATGATGAT 356  
QY 312 TATGCAAAATGCTTTTATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 371  
DB 357 TATGCAAAATGATTTTATCCAGAGAACTGAGATTTG-AGGTCAACGCTTTAAATATC 415  
QY 372 TTGAATGTTCTGAGCTTGAATTAACCATT-AGTGTCTTATTCAGATGATGATGAT 430  
DB 416 CTGGATATTCCTGATCTTGAAGATTAACATATGGAATATTTATTCAGATGATGAT 475

QY 431 TATTAGCACTATATTTTCTCTGTGACAGAAATTCACGTGAGGAAATCAACATCA 490  
DB 476 TATTAGCACTATATTTTCTCTGTGACAGAAATTCACGTGAGGAAATCAACATCA 535  
QY 491 ATGCCGCAAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 550  
DB 536 ATGCAGCAAAATTTTAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 595  
QY 551 TGGGAAAGAGATTTTCAAAAACTTTCAATGATTAAGGCTAAATCATTAAGAGA 610  
DB 596 TGGGAAAGAGATTTTCAAAAACTTTCAATGATTAAGGCTAAATCATTAAGAGA 655  
QY 611 ATTAATATTGAGAGAAAAAATTTCTTAACCTTTCTTCAACCGAGAGATTAACCTG 670  
DB 656 ATTAATATTGAGAGAAAAAATTTCTTAACCTTTCTTCAACCGAGAGATTAACCTG 715  
QY 671 ATGCAGAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 717  
DB 716 ATGCAGAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 762

## RESULT 6

US-08-569-168-1  
; Sequence 1, Application US/08569168  
; Patent No. 5972687  
; GENERAL INFORMATION:  
; APPLICANT: Smigielski, Adam J.  
; APPLICANT: Khurshid, Raymond J.  
; TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lowe, Price, Leblanc & Becker  
; STREET: 99 Canal Center Plaza, Suite 300  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/569,168  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Price, Robert L.  
; REGISTRATION NUMBER: 22,685  
; REFERENCE/DOCKET NUMBER: 1451-015  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-684-1111  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 837 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-569-168-1

Query Match 51.7%; Score 521.4; DB 2; Length 837;  
Best Local Similarity 85.4%; Pred. No. 4,5e-121;  
Matches 604; Conservative 0; Mismatches 101; Indels 2; Gaps 2;  
QY 12 ATTAACACCTGATGATAGATGATATCCACCGTTGAAGCAAAATAGCAGAGATAT 71  
DB 57 AGTAACGCTGATGATGATGATGATATCAACCGTTGAAGCAAAATAGCAGAGATAT 116  
QY 72 AGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 131  
DB 117 AATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 176

QY 132 ATATCGAGCTAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 191  
DB 177 ATATCGAGCTAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 236  
QY 192 ATCTAACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251  
DB 237 CTCTCAACTTCAAGAGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 296  
QY 252 ATATATGCTAATCTTTCATCTGACATTAATTAATTAATTAATTAATTAATTAATTAAT 311  
DB 297 ATATATGCTAATCTTTCATCTGACATTAATTAATTAATTAATTAATTAATTAATTAAT 356  
QY 312 TATGCAAAATGATTTTAAAGAAATGAATGATTAATTAATTAATTAATTAATTAAT 371  
DB 357 TATGCAAAATGATTTTAAAGAAATGAATGATTAATTAATTAATTAATTAATTAAT 415  
QY 372 TTGAAATGTTCTGAGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 430  
DB 416 CTGGAATATTCCTGATCTTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 475  
QY 431 TATTAGCACTATATTTTCTCTGTGACAGAAATTCACGTGAGGAAATCAACATCA 490  
DB 476 TATTAGCACTATATTTTCTCTGTGACAGAAATTCACGTGAGGAAATCAACATCA 535  
QY 491 ATGCCGCAAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 550  
DB 536 ATGCAGCAAAATTTTAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 595  
QY 551 TGGGAAAGAGATTTTCAAAAACTTTCAATGATTAAGGCTAAATCATTAAGAGA 610  
DB 596 TGGGAAAGAGATTTTCAAAAACTTTCAATGATTAAGGCTAAATCATTAAGAGA 655  
QY 611 ATTAATATTGAGAGAAAAAATTTCTTAACCTTTCTTCAACCGAGAGATTAACCTG 670  
DB 656 ATTAATATTGAGAGAAAAAATTTCTTAACCTTTCTTCAACCGAGAGATTAACCTG 715  
QY 671 ATGCAGAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 717  
DB 716 ATGCAGAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 762

## RESULT 7

US-09-806-708B-22/c  
; Sequence 22, Application US/09806708B  
; Patent No. 6784342  
; GENERAL INFORMATION:  
; APPLICANT: The University of British Columbia  
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
; FILE REFERENCE: 4810-58741  
; CURRENT FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: US 60/147,133  
; PRIOR FILING DATE: 1999-08-04  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 22  
; LENGTH: 1141  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; NAME/KEY: Promoter  
; LOCATION: (1)-(1141)  
; OTHER INFORMATION: consensus sequence of A.T., L.A., and B.N. FARI promoters  
US-09-806-708B-22

Query Match 6.2%; Score 62.6; DB 4; Length 1141;  
Best Local Similarity 11.1%; Pred. No. 2.3e-06;  
Matches 81; Conservative 280; Mismatches 360; Indels 6; Gaps 1;  
QY 185 ATATGATCTTAACCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 244  
DB 834 ACRNRTWAWKSKWNN 775

[illegible]

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300-6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELERX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpc-F18
US-08-232-463-14

Query Match      5.9%; Score 59.8; DB 1; Length 7218;
Best Local Similarity 6.0%; Pred.No. 2.2e-05;
Matches 25; Conservative 225; Mismatches 167; Indels 0; Gaps 0;

QY   TCTATTAACTCTATNCGTCTGTAACCTTCAGTCGGAGAAGGATTTTTCAAAATACTTT    580
DB   TGTAATTACTATCTATGTGAAGTAGTTAAGAGATAGAGAATTTGGTACRRRRRRRRR    1425
QY   ACAATGATGATGAGGCTAAATCATTTAGAGATTATATTGAGAGAAAAACTTTCTAAC    640
DB   RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR    1365
QY   CTTTCTTGCAACCACCGCAGAGATTACTGTATGGCAGATATGTTATTGGCTGCACAA    700
DB   RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR    1305
QY   CAGAAAGGCCAAATVGAGAGTAGTTTAAAGAACTTAAAAATAACAATCTAGAAATG    760
DB   RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR    1245
QY   GATTTCATATATGAAAGGCGCTCAAAAATAAGTATATGTTTATAAAGAGGTAC    820
DB   RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR    1185
QY   AAAAGGTACGCTCCACAGACAGCAGCGAAGTATTGTGACGCCAGTGCGACTAAC    880
DB   RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR    1125
QY   TGGAAAAATTCGCAATATATTATATAGTGTAGAGCTAAGCCAAAATAACAGCGGTA    937
DB   RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR    1068

RESULT 9
US-09-806-708B-23
; Sequence 23, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
```



```
QY 824 AGGTAAGCTCCACAGACGAGGAAAGTATTGCTACAGCCAGTGCAGTAACCTCG 883
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 775 GAYTCTTTNNNTNNNTAARDGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 833
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 884 AAAAATGCGCATATTATATATAGTGTAGGCTAAGCCAAAAGACAGGTACCTTTA 943
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 834 TNNNNNNNNNNNNAVMMWTKYTTTDDRMFBATYNNNNNNNNNNNNNNNNNNNN 893
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 944 CTCAAAATGATAGTACATACATACATACATACATACATACATACATACATACAT 978
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 894 AWMKWDATFMNNATYNNNGTAMRTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 928

RESULT 11
US-09-601-198-153
; Sequence 153, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Casseil, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glaes, Jennifer S.
; APPLICANT: Glaes, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; TITLE OF INVENTION: UREAPLITICUM
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 153
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-153

Query Match      4.8%; Score 48.8; DB 4; Length 1866;
Best Local Similarity 47.3%; Pred. No. 0.0076;
Matches 249; Conservative 0; Mismatches 267; Indels 10; Gaps 3;

QY 122 ATGGAATTGAATATGAGCTAAGAAATAATATATATACCTTACGCTTGTAGTGTGA 181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1037 ATGAAAAAACAACAAATATATATATATATATATATATATATATATATATATAT 1096
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 TTCATATGTATCTAACTTCTGATGATATATATATATATATATATATATATATAT 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1097 TTCAAAATATATATGATCAATATATATATATATATATATATATATATATATAT 1156
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 -ATTATACAGATATATATATCTTCAATCTTCACTATATATATATATATATATAT 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1157 ATATTACCAAAAATATATATATATATATATATATATATATATATATATATATAT 1216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 299 AAATTTCTAAGATATGCAAAATGTTTATATATATATATATATATATATATATAT 358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1217 AAATTTCAAAAATATCTAATTTCAAAAATGATATATATATATATATATATATAT 1276
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 ATCCCAAAACATTTGGATGTTCTGACCTGAAATATATATATATATATATATATAT 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1277 AATTAATGAAATTTGATATATATATATATATATATATATATATATATATATAT 1336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 419 ATGACGATTAATTT-----ATTAGCACTATATATTTTCTGTAAGGAATTCACATG 473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1337 ATTAGCAAAAATATATATATATATATATATATATATATATATATATATATATAT 1396
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 474 GGAATAATCAACATCAATGCGCAAGATTTTAAATTAATTAATTAATTAATTAATTA 533
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1397 TGAATTTAAAGCA--CTTGAATTAAGTTTATATATATATATATATATATATATTA 1454
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 534 ATCTGCTGTAACCTTCACTGGAAGGAGATTTTCAAAAACCTTTTACATATGATTTAGA 593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 1455 AATCTCTTACTAAGAAGATGTTTAGACCAAGTCAAAAATATTTAAAAAGTTAAAAA 1514
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 594 GCGTAATCATTTAGAAATTAATTAATTTAGAGAAAAAATCTTCTTAA 639
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1515 ACTTAACCATATATGATTTATATATTTTCAATTAATTAATTTAGAA 1560
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-134-001C-1161
; Sequence 1161, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1161
; LENGTH: 729
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1161

Query Match      4.7%; Score 47.8; DB 3; Length 729;
Best Local Similarity 52.2%; Pred. No. 0.0097;
Matches 106; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 206 ATGACTATTATTAAGATTAAGAAGACTGCTGAGAGATTTATCAAGATATATGTTAATC 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 476 ATGCTATCATATATATATCAAAACCTGATGATTTGAAGATTTATATATATGATTTATA 535
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 TTTCACTGCACTATTAAGTGAAGAAATGTTCAAAATTTCTAAGATATGCAAAATGTT 325
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 536 ATAAATTAATTAATTAATCAATTCATTTGATGATTCAGTTTCAAGAGATGATTAACGATA 595
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 326 TTTATTAAGAATGAATGATTTTGAAGTCAATATCCCAAAAATTTGAATGTTCCG 385
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 596 GTTATATGCTGATTAATGAGTTTGATATCCAACTGATATTAACAAATATTTTATTAACG 655
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 386 AGCTTGAATAATTAACCATTTAGT 408
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 ACTAGCTAAGATTTATATAGT 678
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-806-708B-23/C
; Sequence 23, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 23
; LENGTH: 1055
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)-(1055)
; OTHER INFORMATION: consensus sequence of A.T. and L.A. FAEI promoters
US-09-806-708B-23
```

Query Match 4.7%; Score 47.6; DB 4; Length 1055;  
Best Local Similarity 20.3%; Pred. No. 0.012;  
Matches 154; Conservative 176; Mismatches 429; Indels 1; Gaps 1;

```
QY 12 ATTAACACCTGATGATAGAGGATATCCACCCCTGAAAAGCAATAGCAGAGATAT 71
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 763 ATAAATWGTGTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 72 AGTACGTACTAATCTTAAGCAACAGAGAGGAGCATACAGATCATATGAAATGA 131
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 703 TTTTCTCTTAATATATATATATATATATATATATATATATATATATATATAT
QY 132 ATATGACCTAAGAAAATATATATATAGCTTACGCTTGTGCTGATGATATCATATGT 191
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 643 TTAATATGCAATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 192 ATCTAACTTCTGATGATCTATATATAGAAATGAAGAGCTGCTGAGAAATTTTCAAGA 251
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 583 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 252 ATATATGCTATATCTTATCTGATCTGATGCTGAAATGCTGATCAATTTTCAAGA 311
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 523 AMRTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 312 TATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 463 WAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 372 TTGGAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 403 WMTTAAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 432 ATTAGCATATATATATATATATATATATATATATATATATATATATATATATAT 491
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 343 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 492 TGCCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 551
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 283 SAMAMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 552 GGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 611
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 223 NKATTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 612 TTAATGAGGAAAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 671
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 163 TWTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 672 TGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 731
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 104 MTYAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 732 AGAATCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 44 MCANNTGTRWMAACAACMAAAYCANNMWCRTMRKXTTMMT 5
```

RESULT 14  
US-08-956-171E-531  
Sequence 531, Application US/08956171E  
Patent No. 6593114

GENERAL INFORMATION:  
APPLICANT: Charles Kunsch

Gil H. Choi

Patrick S. Dillon

Craig A. Rosen

Steven C. Barash

Michael R. Ransom

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 531:

SEQUENCE CHARACTERISTICS:

LENGTH: 942 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 531:

US-08-956-171E-531  
Query Match 4.7%; Score 47.4; DB 4; Length 942;  
Best Local Similarity 47.8%; Pred. No. 0.013;  
Matches 138; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

```
QY 66 AGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 125
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 313 AGAAAAGCTATCTTAACGCAATGAGCAATTTGATGAAACAGATTAAGCTTTCTAT 372
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 126 AATGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 373 AGATGATATTAAGATCTTAACCTAGTATGATGATGATGATGATGATGATGATGAT
QY 186 TAAATGATCTTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 245
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 433 CATGAGGAGAAATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 246 TCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 305
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 493 AACCCCAATCAATGCTATTTTATATTTTATGATTAATGATGATGATGATGATGAT
QY 306 TAAAGATATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 553 AAAACAAAGAACATACGCTTTTATTAAGAAAGAGATTTATTAAGAT 601
```

RESULT 15  
US-08-781-986A-531  
Sequence 531, Application US/08781986A  
Patent No. 6737248

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5255

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

```

; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 531:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 942 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-781-986A-531

```

Query Match 4.7%; Score 47.4; DB 4; Length 942;

Best Local Similarity 47.8%; Pred. No. 0.013; Mismatches 151; Indels 0; Gaps 0;

```

Matches 138; Conservative 0;
;
QY 66 AGATATAGTACGTATATCTTAACCTTAAGCAACAGATGAGGTCATACACATCATATG 125
DB 313 AGAAAAAGTACTTAAGCAATGAGCAATTAAGTGAACAGAAATTAAGTCACTTCTAT 372
QY 126 AATTGATATCGAGCTAAGAAATTAATTAAGCTTACGCTTGGCTGTAAGTGATTTCA 185
DB 373 AGATGATATTAAGAGTCTAACTAGTATCGTATATTGATGATGGTGCAACTGGTCTGG 432
QY 186 TAATGATATTAACCTCCGATGACTATATTAAGAAATAAGAGACTGCTGAGAAATTTA 245
DB 433 CATAGAGAGAAATTAATATGATGACATATGCGATGAGAGATATTTTAAAGAGACGA 492
QY 246 TCAAGAAATATATGCTATATCTTCAATCTGCACTATTAAGTGAAATGATCAAAATTC 305
DB 493 AACGCATCAATGCGATTTTGTATTTTAAATGATGATGATGATGATGATGATGATGAT 552
QY 306 TAAAGATATGCAAAATGTTTTTATTAAGATGAATGAACTGATTTTGAAGGT 354
DB 553 AAAACAAGAGACATACGCTTTATTAAGAAAGATCTTATTAAGAT 601

```

Search completed: November 21, 2004, 08:44:25  
 Job time : 122 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 06:27:26 ; Search time 4336 Seconds  
(without alignments)  
10993.546 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1008

Sequence: 1 atggttatcatatcaacc.....ctcatataaataatataga 1008

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ha: \*  
2: gb\_hgc: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_to: \*  
11: gb\_sbs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vt: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	100.0	1008	AR407888	Sequence
2	1008	100.0	1008	AX029370	Sequence
3	1008	100.0	1388	AX029374	Sequence
4	1007	99.9	1007	BD080162	Toxin gen
5	992	98.4	349652	BX571866	Photocorhab
6	992	98.4	349980	AX770906	Sequence
7	663.8	65.9	1107	AR407887	Sequence
8	663.8	65.9	1107	AX029369	Sequence
9	663.8	65.9	1107	BD080161	Toxin gen
10	663.8	65.9	1205	AR407889	Sequence
11	663.8	65.9	1205	AX029373	Sequence
12	628.2	62.3	1272	AR082011	Sequence
13	521.4	51.7	834	AR082008	Sequence
14	521.4	51.7	837	AR082007	Sequence
15	62.6	6.2	1141	AX083744	Sequence
16	62.2	6.2	11000	AC116984_3	Continuation (4 of
17	61.2	6.1	5658	AB102780	Dicystose
18	60.6	6.0	149783	CR450800	Danio rer
19	60.6	6.0	330199	BX901880	Danio rer

20	60	6.0	236120	14	AF063866	AF063866 Melanopl
21	59.8	5.9	7218	6	166494	166494 Sequence 14
22	59	5.9	34347	2	BX957360	BX957360 Danio rer
23	58.4	5.8	178247	5	BX248504	BX248504 Zebrafish
24	58.4	5.8	335050	3	PPA929356	PPA929356 Plasmodu
25	58	5.8	191904	9	AC005699	AC005699 Homo sapi
26	57.8	5.7	6717	1	AF211135	AF211135 Carbone11
27	57.8	5.7	7291	1	AF211133	AF211133 Carbone11
28	57.8	5.7	258658	3	AE014832	AE014832 Plasmodu
29	57.6	5.7	252394	3	AE014833	AE014833 Plasmodu
30	57.2	5.7	600	3	AF201316	AF201316 Dugesia p
31	57.2	5.7	86826	3	PFMAL3P5	PFMAL3P5 Plasmodu
32	57	5.7	110000	2	PFMAL13_07	PFMAL13_07 Plasmodu
33	57	5.7	250029	3	AE014820	AE014820 Continuation (8 of
34	57	5.7	341050	3	PPA929357	PPA929357 Plasmodu
35	56.8	5.6	110000	2	PFMAL8P1_11	PFMAL8P1_11 Continuation (12 o
36	56.8	5.6	253924	3	AE014822	AE014822 Plasmodu
37	56.4	5.6	185343	2	BX248402	BX248402 Danio rer
38	56	5.6	115985	5	BX890614	BX890614 Zebrafish
39	56	5.6	143342	2	CR388191	CR388191 Danio rer
40	55.8	5.5	110000	3	AC116984_4	Continuation (5 of
41	55.6	5.5	914	3	PFRESA	X55124 P.falciparu
42	55.6	5.5	180874	2	BX569779	BX569779 Danio rer
43	55.4	5.5	49306	3	AC115606	AC115606 Dicystose
44	55.2	5.5	84618	5	AL591365	AL591365 Zebrafish
45	55.2	5.5	85916	3	AC117080	AC117080 Dicystose

#### ALIGNMENTS

RESULT 1	AR407888	Sequence 2 from patent US 6630619.	DNA	linear	PAT 18-DEC-2003
LOCUS	AR407888	Sequence 2 from patent US 6630619.	DNA	linear	PAT 18-DEC-2003
DEFINITION	AR407888	Sequence 2 from patent US 6630619.	DNA	linear	PAT 18-DEC-2003
ACCESSION	AR407888	Sequence 2 from patent US 6630619.	DNA	linear	PAT 18-DEC-2003
VERSION	AR407888.1	GI:40157856	DNA	linear	PAT 18-DEC-2003
KEYWORDS	Unknown.		DNA	linear	PAT 18-DEC-2003
SOURCE	Unknown.		DNA	linear	PAT 18-DEC-2003
ORGANISM	Unknown.		DNA	linear	PAT 18-DEC-2003
REFERENCE	1 (bases 1 to 1008)		DNA	linear	PAT 18-DEC-2003
AUTHORS	East, P.D.		DNA	linear	PAT 18-DEC-2003
TITLE	Toxin genes from the bacteria Xenorhabdus nematophilus and		DNA	linear	PAT 18-DEC-2003
JOURNAL	photocorhabus luminescens		DNA	linear	PAT 18-DEC-2003
FEATURES	Patent: US 6630619-A 2 07-OCT-2003;		DNA	linear	PAT 18-DEC-2003
source	Location/Qualifiers		DNA	linear	PAT 18-DEC-2003
ORIGIN	1..1008		DNA	linear	PAT 18-DEC-2003
Query Match	100.0%; Score 1008; DB 6; Length 1008;		DNA	linear	PAT 18-DEC-2003
Best Local Similarity	100.0%; Pred. No. 5.2e-171;		DNA	linear	PAT 18-DEC-2003
Matches 1008; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		DNA	linear	PAT 18-DEC-2003
QY	1 ATGTTATACATTAACCTGATGATGAAGTGATATCCACCCGTTGAAAGCAATTA	60			
DB	1 ATGTTATACATTAACCTGATGATGAAGTGATATCCACCCGTTGAAAGCAATTA	60			
QY	61 GCAGGAGATATAGTACGATTAACCTTAAGCAAGATGAGGATCATACAGATCA	120			
DB	61 GCAGGAGATATAGTACGATTAACCTTAAGCAAGATGAGGATCATACAGATCA	120			
QY	121 TATGGAATTAATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	180			
DB	121 TATGGAATTAATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	180			
QY	181 ATTGATATATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG	240			
DB	181 ATTGATATATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG	240			
QY	241 ATTATCAAGATATATATGATTAATGATTAATGATTAATGATTAATGATTAATG	300			

Db 241 ATTTTCAGAGATATATATGCTTAATCTTTATCTGACCTATATAGTGGAATATGCTATCA 300  
Qy 301 ATTTCTAAAGATATGCGCAATATGTTTTTAAGAAATGAATGATTTTGAAGCTCAATAT 360  
Db 301 ATTTCTAAAGATATGCGCAATATGTTTTTAAGAAATGAATGATTTTGAAGCTCAATAT 360  
Qy 361 CCTCAAAACATTGGAAATGTTCTCTGAGCTTGAATAAACCATTGAGTCTTATTCAGAT 420  
Db 361 CCTCAAAACATTGGAAATGTTCTCTGAGCTTGAATAAACCATTGAGTCTTATTCAGAT 420  
Qy 421 GACGATTAATATTATGACCTATATTTTCTCTGTACAGAAATTCACCTGAGGAAAT 480  
Db 421 GACGATTAATATTATGACCTATATTTTCTCTGTACAGAAATTCACCTGAGGAAAT 480  
Qy 481 CAACATCAAAATGCGCAGAAATTTTAAATTAATGATTTCTTATTTACCTTATCTGCT 540  
Db 481 CAACATCAAAATGCGCAGAAATTTTAAATTAATGATTTCTTATTTACCTTATCTGCT 540  
Qy 541 GTAACCTTCACTGGGAAAGAGATTTTTCAAAAAACCCTTTACAAATGATTAAGGCTTAA 600  
Db 541 GTAACCTTCACTGGGAAAGAGATTTTTCAAAAAACCCTTTACAAATGATTAAGGCTTAA 600  
Qy 601 TCATTAGAGAAATTAATGAGAGAAAAAACCCTTCTTGAACCGACCGAG 660  
Db 601 TCATTAGAGAAATTAATGAGAGAAAAAACCCTTCTTGAACCGACCGAG 660  
Qy 661 AGATTACCTGATGCGAGAAATAGGTTATTTGCTGACCAAGAACGCGCTTAATGAGA 720  
Db 661 AGATTACCTGATGCGAGAAATAGGTTATTTGCTGACCAAGAACGCGCTTAATGAGA 720  
Qy 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTCTAATATGAGAGG 780  
Db 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTCTAATATGAGAGG 780  
Qy 781 GCTGCAAAACAAAAGTATAGTTCATTTTAAAGAGGTACAAAAGGCTACCTCCACAG 840  
Db 781 GCTGCAAAACAAAAGTATAGTTCATTTTAAAGAGGTACAAAAGGCTACCTCCACAG 840  
Qy 841 ACAGCAGGGAAGATATGTTGTAACAGCCAGTGCAGTAACTTGAAAAATTTGCCGATAT 900  
Db 841 ACAGCAGGGAAGATATGTTGTAACAGCCAGTGCAGTAACTTGAAAAATTTGCCGATAT 900  
Qy 901 TTATATAGTGTAGAGCTTAAGCCAAAAAGACAGGTTAACCCTTACTCAAAATGATCTGAC 960  
Db 901 TTATATAGTGTAGAGCTTAAGCCAAAAAGACAGGTTAACCCTTACTCAAAATGATCTGAC 960  
Qy 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATATGA 1008  
Db 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATATGA 1008

RESULT 2  
AX029370 1008 bp DNA linear PAT 16-SEP-2000  
LOCUS Sequence 2 from Patent WO903328.  
AX029370  
ACCESSION AX029370.1 GI:10190184  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Photorhabdus luminescens  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Photorhabdus.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMUN SCIENT IND RES ORG (AU); EAST PETER DAVID (AU)  
FEATURES  
source  
1..1008  
/organism="Photorhabdus luminescens"  
/mol\_type="unassigned DNA"

ORIGIN /db\_xref="taxon:29488"  
Query Match 100.0%; Score 1008; DB 6; Length 1008;  
Best Local Similarity 100.0%; Pred. No. 5.2e-171;  
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGTTTATACATTTAACCCTGATGATGAAAGTGATATCCACCCGTTGAAAAAGCAATA 60  
Db 1 ATGTTTATACATTTAACCCTGATGATGAAAGTGATATCCACCCGTTGAAAAAGCAATA 60  
Qy 61 GAGAGATATATAGTACGATATCTAACTTTAAGCAACAGATGAGGTCATACAGATCA 120  
Db 61 GAGAGATATATAGTACGATATCTAACTTTAAGCAACAGATGAGGTCATACAGATCA 120  
Qy 121 TATGGAATTGAAATTCGAGCTTAAGAAATAATATTAGCTTACGCTTGGCTGAAGTGT 180  
Db 121 TATGGAATTGAAATTCGAGCTTAAGAAATAATATTAGCTTACGCTTGGCTGAAGTGT 180  
Qy 181 ATTCATTAATGTAATCTAAACCTTCTGATGATCTATTATTAAGAAATTAAGAGACTGCTGAGAG 240  
Db 181 ATTCATTAATGTAATCTAAACCTTCTGATGATCTATTATTAAGAAATTAAGAGACTGCTGAGAG 240  
Qy 241 ATTTATCAAGAAATATATGCTTAATCTTCACTGACCTATTAGGTAAGTGAATCA 300  
Db 241 ATTTATCAAGAAATATATGCTTAATCTTCACTGACCTATTAGGTAAGTGAATCA 300  
Qy 301 ATTTCTAAAGATATGCGCAAAATGTTTTTAAAGATGAATCTGATTTTGAAGTCAATAT 360  
Db 301 ATTTCTAAAGATATGCGCAAAATGTTTTTAAAGATGAATCTGATTTTGAAGTCAATAT 360  
Qy 361 CCTCAAAACATTGGAAATGTTCTCTGAGCTTGAATAAACCATTGAGTCTTATTCAGAT 420  
Db 361 CCTCAAAACATTGGAAATGTTCTCTGAGCTTGAATAAACCATTGAGTCTTATTCAGAT 420  
Qy 421 GACGATTAATATTATGACCTATATTTTCTCTGTACAGAAATTCACCTGAGGAAAT 480  
Db 421 GACGATTAATATTATGACCTATATTTTCTCTGTACAGAAATTCACCTGAGGAAAT 480  
Qy 481 CAACATCAAAATGCGCAGAAATTTTAAATTAATGATTTCTTATTTACCTTATCTGCT 540  
Db 481 CAACATCAAAATGCGCAGAAATTTTAAATTAATGATTTCTTATTTACCTTATCTGCT 540  
Qy 541 GTAACCTTCACTGGGAAAGAGATTTTTCAAAAAACCCTTTCAATGATTTAGAGCTTAA 600  
Db 541 GTAACCTTCACTGGGAAAGAGATTTTTCAAAAAACCCTTTCAATGATTTAGAGCTTAA 600  
Qy 601 TCATTAGAGAAATTAATGAGAGAAAAAACCCTTCTAATCCTTCTTGCACACCGAG 660  
Db 601 TCATTAGAGAAATTAATGAGAGAAAAAACCCTTCTAATCCTTCTTGCACACCGAG 660  
Qy 661 AGATTACCTGATGCGAGAAATAGTATTTGCTGACCAAGAGCGCTTAATGAGA 720  
Db 661 AGATTACCTGATGCGAGAAATAGTATTTGCTGACCAAGAGCGCTTAATGAGA 720  
Qy 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780  
Db 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780  
Qy 781 GCTGCAAAACAAAAGTATAGTTCATTTTAAAGAGGTACAAAAGGCTACCTCCACAG 840  
Db 781 GCTGCAAAACAAAAGTATAGTTCATTTTAAAGAGGTACAAAAGGCTACCTCCACAG 840  
Qy 841 ACAGCAGGGAAGATATGTTGTAACAGCCAGTGCAGTAACTTGAAAAATTTGCCGATAT 900  
Db 841 ACAGCAGGGAAGATATGTTGTAACAGCCAGTGCAGTAACTTGAAAAATTTGCCGATAT 900  
Qy 901 TTATATAGTGTAGAGCTTAAGCCAAAAAGACAGGTTAACCCTTACTCAAAATGATCTGAC 960  
Db 901 TTATATAGTGTAGAGCTTAAGCCAAAAAGACAGGTTAACCCTTACTCAAAATGATCTGAC 960  
Qy 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATATGA 1008  
Db 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATATGA 1008

Db	772	TCATTAGAGAAATTATATTGAGAGAAAAAACTTCTTAAACCTTTCTTTCGACACCCGAG	831
Qy	661	AGATTACCTGATGGCAAGATAGTTATTGGCTGGACCAACAGAGGGCCTAATATGAGA	720
Db	832	AGATTACCTGATGGCAAGATAGTTATTGGCTGGACCAACAGAGGGCCTAATATGAGA	891
Qy	721	GTCGATTTTAAAGAACTTAAAAATACAAATCTAGAAATGGATTCTTAATATGGAAGG	780
Db	892	GTCGATTTTAAAGAACTTAAAAATACAAATCTAGAAATGGATTCTTAATATGGAAGG	951
Qy	781	GCTGCAAAACAAAAGTATAGTTCATTATATAAAGAGTACAAAAGGTAACGCTCCACAG	840
Db	952	GCTGCAAAACAAAAGTATAGTTCATTATATAAAGAGTACAAAAGGTAACGCTCCACAG	1011
Qy	841	ACAGCAGCGAAAGATATTGGTACAGCAGGAGGACGTAACTCGAAAAATTCGCCAATAT	900
Db	1012	ACAGCAGCGAAAGATATTGGTACAGCAGGAGGACGTAACTCGAAAAATTCGCCAATAT	1071
Qy	901	TTATATATGATGAGCGCTTAAGCCAAAAGACAGGGTAACTTTACTCAAAATGATATGAC	960
Db	1072	TTATATATGATGAGCGCTTAAGCCAAAAGACAGGGTAACTTTACTCAAAATGATATGAC	1131
Qy	961	AATTCATATGACGCTTCACTAGTGTGGAACTCATTTATATAATATATGA	1008
Db	1132	AATTCATATGACGCTTCACTAGTGTGGAACTCATTTATATAATATATGA	1179
RESULT 4			
BD080162			
LOCUS	BD080162	1007 bp	DNA
DEFINITION	Toxin genes from bacteria Xenorhabdus nematophilus and Photorhabdus luminescens.		
ACCESSION	BD080162		
VERSION	BD080162.1	GI:22625765	
KEYWORDS	JP 2001510022-A/2.		
SOURCE	Photorhabdus luminescens		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.		
REFERENCE	1. (bases 1 to 1007)		
AUTHORS	East, P. D.		
TITLE	Toxin genes from bacteria Xenorhabdus nematophilus and Photorhabdus luminescens		
JOURNAL	Patent: JP 2001510022-A 2 31-JUL-2001;		
COMMENT	COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION		
	OS Photorhabdus luminescens		
	PN JP 2001510022-A/2		
	PF 17-JUL-1998 JP 2000502652		
	PR 17-JUL-1997 AU		
	PI PETER DAVID EAST		
	PC C12N15/09, A01H5/00, A01N63/00, C07K14/24, C12N1/10, C12N1/19 PC		
	C12N1/21, C12N5/10.		
	PC C12N7/00, C12N15/00, C12N5/00		
	CC Toxin genes from bacteria Xenorhabdus nematophilus and CC		
	Photorhabdus		
	luminescens		
FEATURES			
Source	CC luminescens		
	FH Key		
	FT 1..1007		
	FT Location/Qualifiers		
	1..1007		
	/organism="Photorhabdus luminescens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:29488"		
ORIGIN			
	99.9%; Score 1007; DB 6; Length 1007;		
	Best Local Similarity 100.0%; Pred. No. 7.8e-171; Indels 0; Gaps 0;		
	Matches 1007; Conservative 0; Mismatches 0;		
Qy	1	ATGCTTATACAAATTACACCTGATGATAGAAAGTGGATATCCACCCGTTGAAAGCAATA	60

Db 1 ATGTTTATCAATTAACACCTGATAGTAGAGTGAATTCACCCGTTGAAAAACAAATA 60  
Qy 61 GCAGAGATATATAGTACGTATACCTTAAGCAAAACAGATGAGGGTCATACAGCATCA 120  
Db 61 GCAGAGATATATAGTACGTATACCTTAAGCAAAACAGATGAGGGTCATACAGCATCA 120  
Qy 121 TATGGAATTTGAATTCGAGCTTAAGAAATTAATTAATTAAGCTTACGCTTGGCTGAAGTGT 180  
Db 121 TATGGAATTTGAATTCGAGCTTAAGAAATTAATTAATTAAGCTTACGCTTGGCTGAAGTGT 180  
Qy 181 ATTCAATATGTATCTAACTTCCTGATGACTATTAATTAAGAAATTAAGAGCTCTGAGAGA 240  
Db 181 ATTCAATATGTATCTAACTTCCTGATGACTATTAATTAAGAAATTAAGAGCTCTGAGAGA 240  
Qy 241 ATTATCAAGATATATATGTCTATATCTTTCATCTGCACTATTAGTGTAATGATGATCA 300  
Db 241 ATTATCAAGATATATATGTCTATATCTTTCATCTGCACTATTAGTGTAATGATGATCA 300  
Qy 301 ATTTCTAAAGATATGCGCAATGCTTTTAAAGAAATGAATGAACTGATTTGAAAGTCAATAT 360  
Db 301 ATTTCTAAAGATATGCGCAATGCTTTTAAAGAAATGAATGAACTGATTTGAAAGTCAATAT 360  
Qy 361 CCTCAAAACATTTGGAATGCTCTGAGCTTGAATAATTAACCAATGAGTCTTATTCAGAT 420  
Db 361 CCTCAAAACATTTGGAATGCTCTGAGCTTGAATAATTAACCAATGAGTCTTATTCAGAT 420  
Qy 421 GACGATTAATTAATTAAGCACTATATTTTCTCTGTAACAGAAATTCACCTGAGAGAAAT 480  
Db 421 GACGATTAATTAATTAAGCACTATATTTTCTCTGTAACAGAAATTCACCTGAGAGAAAT 480  
Qy 481 CAACATCAAAATGCGGCAAGATTTTAAATTAATTAATTAATTAATTAATTAATTCGT 540  
Db 481 CAACATCAAAATGCGGCAAGATTTTAAATTAATTAATTAATTAATTAATTAATTCGT 540  
Qy 541 GTPAATCTACCTGGAAGAGAGATTTTTCAAAAAATTTTAAAGATTAAGAGGCTTAA 600  
Db 541 GTPAATCTACCTGGAAGAGAGATTTTTCAAAAAATTTTAAAGATTAAGAGGCTTAA 600  
Qy 601 TCATTAAGAAATTAATTAAGAGAGAAATTTTCTTAACTTTCTTTCGACACCGCAG 660  
Db 601 TCATTAAGAAATTAATTAAGAGAGAAATTTTCTTAACTTTCTTTCGACACCGCAG 660  
Qy 661 AGATTACCTGATGCGAATAGGTTATTTGCTGCAACAAAGCGCTTAAATGAGAGA 720  
Db 661 AGATTACCTGATGCGAATAGGTTATTTGCTGCAACAAAGCGCTTAAATGAGAGA 720  
Qy 721 GTGAGTTTAAAGACTTAATAATTAACAATCTAGGAATGAAATTTCTAATTAAGAGGG 780  
Db 721 GTGAGTTTAAAGACTTAATAATTAACAATCTAGGAATGAAATTTCTAATTAAGAGGG 780  
Qy 781 GCTGCAAAACAAAGATATAGTCAATTTATTAAGAGGTAACAAAGGGTAACGCTCCACAG 840  
Db 781 GCTGCAAAACAAAGATATAGTCAATTTATTAAGAGGTAACAAAGGGTAACGCTCCACAG 840  
Qy 841 ACAGCAGCGAAAGATTTGCTACAGCAGCTGCGAGTAACCTGAGAAAATTTGCCGAATAT 900  
Db 841 ACAGCAGCGAAAGATTTGCTACAGCAGCTGCGAGTAACCTGAGAAAATTTGCCGAATAT 900  
Qy 901 TTATATAGTGAAGGCTAGCAAAAGACAGGGTAACTTTAATCTCAAAATGATATCTGAC 960  
Db 901 TTATATAGTGAAGGCTAGCAAAAGACAGGGTAACTTTAATCTCAAAATGATATCTGAC 960  
Qy 961 AATACATGACGGTTCATAGTGTGGAACCTTATTAATAATAATATATNG 1007  
Db 961 AATACATGACGGTTCATAGTGTGGAACCTTATTAATAATAATATATNG 1007

RESULT 5  
LOCUS BX571866 349652 bp DNA linear BCT 26-SEP-2003  
DEFINITION Photorhabdus luminescens subsp. laumondii TT01 complete genome;  
ACCESSION BX571866 BX470251

VERSION BX571866.1 GI:36785405  
KEYWORDS complete genome.  
SOURCE Photorhabdus luminescens subsp. laumondii TT01  
ORGANISM Photorhabdus luminescens subsp. laumondii TT01  
Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.  
REFERENCE 1  
Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Taouric, S., Bocs, S., Boursaux-Bude, C., Chandler, M., Dassa, E., Deroose, R., Deruelle, S., Freyssinet, G., Gaudreau, S., Givaudan, A., Glaeser, P., Medigue, C., Lanois, A., Powell, K., Signier, P., Wingate, V., Zouine, M., Boemare, N., Danchin, A. and Kunst, F.  
Complete genome sequence of the endonapathogenic bacterium Photorhabdus luminescens  
Nat. Biotechnol. 11 (1) (2003) In press  
2  
Duchaud, E., Frangeul, L., Rusniok, C. and Kunst, F.  
Direct Submission  
Submitted (23-Apr-2003) L. Frangeul, Institut Pasteur, Genopole, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: lfrangeul@pasteur.fr, fkunst@pasteur.fr  
Location/Qualifiers  
1. 349652  
/organism="Photorhabdus luminescens subsp. laumondii TT01"  
/mol\_type="genomic DNA"  
/strain="TT01"  
/db\_xref="taxon:243265"  
572. .2086  
/locus\_tag="plu2059"  
572. .577  
/locus\_tag="plu2059"  
584. .2086  
/locus\_tag="plu2059"  
/function="Unknown"  
/note="unnamed protein product; Highly similar to unknown protein of Photorhabdus"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAE14352.1"  
/db\_xref="GI:36785406"  
/translation="MSILTPDIVSGITVLEACPVIITVYSGDDTHKVTETDIWYSH EKNTFPDTPRIILDVENNTCTAKFTVYTAGSDADGFTTTPSMDKATDPDAELGAFHS IKDKDFSHAEPEPIIIGDSLTKDLSPVADDERPSPFNSGLINITYALSKTBPPLFYFO IPLVIDNLVRIEYDGENITTEILPEIKIINNKKYIINTDQKGVALKPFRKKEKIGN YTVMFILPEFLKXASNIILFIEDTAPALPADIIBELDNRLTPSVAGNSGTGHVSI IPFYKAKPRDFILVIRENDDELGAETVITDLYOLNPLFLVYASVPNIGNEILC YVVDVNVGNAHVAAGNNGIDYDSVINRATVYNGVINEPDKRTLAAPRIFAHKOOLEFEGSRINLE NIIIGGLNAHVAAGNNGIDYDSVINRATVYNGVINEPDKRTLAAPRIFAHKOOLEFEGSRINLE GETIIQLQAPFPFKVGSYCDGTPGSVYITVYSDVTKESIWSGHINTAFVG"  
2442. .3971  
/locus\_tag="plu2060"  
2442. .3971  
/locus\_tag="plu2060"  
/function="Unknown"  
/note="unnamed protein product; Highly similar to unknown protein of Photorhabdus"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAE14353.1"  
/db\_xref="GI:36785407"  
/translation="WFENMRFPNMSILTRADIVSGITVLEACPVIITVYSGDESKVT ETTDIVSDQKVSFEDETPRIISRVKNNICRAKFPVADHSAVANDFTIIFSDKTTDPP AKLEIAFPIIDKNGFSHAEPFMIIGSLVNDLSPVADDEPSEFNGSLIVINYALNK TDDPLPYFOIPLVTDNRVIRFGEBOGTEIEIPFIOPLANSKYIINTNSGFALARA FPGSGEGNVYVNMFTTELQFSRSSNIILFITDIAPIIWMPOIKELDGRILPSVG NSSGTGFHYSIPFYBAGAKRDIRPLMIRENDNEIGSTSVITDIANDLSDSPISAPYASI PNIGDNEILYIVADTMGNVAISSLYFTLTVSVNPQPEIRRTILAAAKIFHKDSEL KORGINLVKILGGGLDARIVAGKNGIDVDSIINVRATVYNGVINEPDKRTLAAPRIFAHKOOLEFEGSRINLE GETIIQLQAPFPFKVGSYCDGTPGSVYITVYSDVTKESIWSGHINTAFVG"  
4296. .5909  
/locus\_tag="plu2061"  
4296. .4302

```

CDS
    /locus_tag="plu2061"
    4311..5909
    /locus_tag="plu2061"
    /function="Unknown"
    /note="unnamed protein product; Highly similar to unknown
    protein of Photorhabdus"
    /codon_start=1
    /transl_table=11
    /protein_id="CAE14354.1"
    /db_xref="GI:36785408"
    /translation="MWKINPDKLDVPEYLLMFENMLNLSLTADIVSGITILEGC
    PVVITVYSGDEHKVTEATDIVYSDQDVSFDEIPRISRDVKNNIYAKVGSAYSD
    ASDPTFMSMDKADPDPAKLEIFHSIGKNDSEYEGFTPIIGDSWVDFPSADNR
    PSFNDRLINIVALKNDKDLPEYQIPLVTDKPYRIGVNRGOIREILPEISEKON
    COYIINDSKGFILVYPPKSEGNYSVMGLTKPYLKSSNITMTDEPTIPA
    PDIAELGNKMLPSVGSNSDTPHYIIPYEGVRPDRVFLITREDDDETQASAVITD
    ILQDMSFISVYANIPNTGNNELYYVADATGVAVASARYFNLTKVLVNPPEIR
    ILAVRIFSHDQVEFRQSSKINLASIIGGIDAHIAVGERNGEINNINVARLYVNG
    IINENPFESKIYSLAHTVADVMDAGEAIIPLGRAIFTNNSYRDGTGMYITVSVC
    MESESTKSHKIWRGIDTVPGR"
    join(5974..5979,5993..6679)
    /locus_tag="plu2062"
    5974..5979
    /locus_tag="plu2062"
    5993..6679
    /locus_tag="plu2062"
    /function="Unknown"
    /note="unnamed protein product; Some similarities with
    unknown protein of Photorhabdus"
    /codon_start=1
    /transl_table=11
    /protein_id="CAE14355.1"
    /db_xref="GI:36785409"
    /translation="WNGISISSVITITENAMSDLDIDYENLNPQNTVTLXGLPGHKRYAT
    TSIGAFIVDDINLDQIILTDETEGKHPYRSPHEKNINSESEFSAFVADQEDI
    NKWMSFPLTFNGYRQSDERAITFTAYNTTGA PADGKTPCSYILFIDREHNDIQIRI
    RVNNNALIDGYNKQWADIPLKEDGSAATVNVISNTYGVKNVWLTPADSDSGKNVFLS
    FRPTPMGEET"
    6664..6669
    /locus_tag="plu2063"
    6664..6669
    /locus_tag="plu2063"
    6676..8001
    /locus_tag="plu2063"
    6676..8001
    /locus_tag="plu2063"
    /function="Unknown"
    /note="unnamed protein product; Some similarities with
    unknown protein of Photorhabdus"
    /codon_start=1
    /transl_table=11
    /protein_id="CAE14356.1"
    /db_xref="GI:36785410"
    /translation="WITNNDNTPISLSSPELPOAHDGLIDVVOONODLIKKVYHE
    WLRIGDMIVVAVGIFESRPYIYONINILDFYTTIPFSEIIPGCVAVTYSADPENEN
    ESKRPVKTINSPIKILPSIYEDTSVTHSSIVGSAVNHSTYQMKQGSBIVT
    HMQGNEDIDPILARAMDAPITVSDEVKQGVSLVVPAAENVLCLAGSEPGGTGYEV
    GNSAPGSVILSLTDMSSAMISIEGAPMSNEYPRIINPNLAVVYBPGRIRJASAV
    ESGSIVGEGGAKYQSPNSFGIAYPHLOSIVAGSRCVTSLEDGSIITGVGOAVSEW
    RDGNNISAVAFSTSEPADGVSCGVIYVDRYVSGRLVLRVTGSAIISGYGNER
    ELQVAKSGDGTPIHTNTKAKKNVILSLPDTDSIQPSINFKHPV"
    8105..9084
    /locus_tag="plu2064"
    8105..9084
    /locus_tag="plu2064"
    8116..9084
    /locus_tag="plu2064"
    /function="Unknown"
    /note="unnamed protein product; Weakly similar to putative
    invasin and adhesin"
    /codon_start=1
    /transl_table=11
    /protein_id="CAE14357.1"

```

```

    /db_xref="GI:36785411"
    /translation="MNDEQAPKEAKTLNMLSEKTEPADRPIDIGVADQGYLLAGV
    NIIFTLNOSMVFNKENDRKIKGTGVLGVSVIISTDSPPCGANYAACLEDTEIE
    APPIQVEFTYAPQSAVALYNLDLEKNNAWTCGADENKRYATVYDLGNGIKNOELTTPA
    DSGATVPSGGMTDKDKRTIASIOSRAGEVYILAQGALEKAIKMPETTPYDQTKI
    EFPAYLVSFSSGIIAGLYKNGVGEPPVTIGVSGSYLSISDQKMTTDDKNGYFSFR
    VSSYSRGVSPPHNVPPMSSEFSSVYISCIDGRVYSQAMIELTPLY"
    9552..9905
    /locus_tag="plu2065"
    9552..9557
    /locus_tag="plu2065"
    9564..9839
    /locus_tag="plu2065"
    /function="Unknown"
    /note="unnamed protein product; Similar to probable
    membrane protein Ychh of Becherichia coli"
    /codon_start=1
    /transl_table=11
    /protein_id="CAE14358.1"
    /db_xref="GI:36785412"
    /translation="MKROSAPFIVGNVLMGVMTFIVSIVYTLPHIFGFGMPFELS
    AALIGVFICALVWLAGAGVGREGVADRYMWRNRYDERYRGRSHRYP"
    9881..9905
    /locus_tag="plu2065"
    complement(9935..10896)
    /gene="ptraA"
    /locus_tag="plu2066"
    complement(935..10882)
    /gene="ptraA"
    /locus_tag="plu2066"
    /codon_start=1
    /transl_table=11
    /product="Ribose-phosphate pyrophosphokinase
    (phosphoribosyl pyrophosphate synthetase)"
    /protein_id="CAE14359.1"
    /db_xref="GI:36785413"
    /translation="MPDMKLFAGNAIPDLAQRVANRLYTNLGAADVGRFSDEYSVQV
    NENVRGGDIFIGSTCAPTNDNLMELVMDALBASAGRTAVIPEYGVRODRVR
    SARVPTRKAVVADPLSSGVDRVLTVDLHAEOLQFPDVPDVFSPILLLEDMLOXE
    LENPIVSPDIGVVRARAIKLANDMDALIDRRPRANVSQVWHIIGVAGDCVL
    VDDMIDVGTGLCKAAEALKERGAARVAVATHPIPSGNAVDNINNSVIDEVVCDTIP
    LSAEIKALKNKVTLTSLGMLAEAIRRIISNEBSISAMPEH"
    complement(10891..10896)
    /gene="ptraA"
    /locus_tag="plu2066"
    complement(10946..11821)
    /gene="ispB"
    /locus_tag="plu2067"
    complement(10946..11821)
    /gene="ispB"

CDS
    Query Match 98.4%; Score 992; DB 1; Length 349652;
    Best Local Similarity 99.0%; Pred. No. 8e-169;
    Matches 998; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

    QY 1 ATGGTTATACATTAACACCTGATGATAGTAGTGATATCCACCGCTGAAGAATA 60
    DB 306840 ATGGTTATACATTAACACCTGATGATAGTAGTGATATCCACCGCTGAAGAATA 306899
    QY 61 GCAGAGATATAGTACGTATCTAACTTAAAGCAACAGATGAGGCTCATACAGATCA 120
    DB 306960 GCAGAGATATAGTACGTATCTAACTTAAAGCAACAGATGAGGCTCATACAGATCA 306959
    QY 121 TATGGAATTAATCGAGCTAAGAAATAATATATAGCTTACGCTTGGCTGAAGTGT 180
    DB 306960 TATGGAATTAATCGAGCTAAGAAATAATATATAGCTTACGCTTGGCTGAAGTGT 307019
    QY 181 ATTGATATGATCTAACTTCCGATGATCATATTATAGATTAAGAATAAGAGACTGAGAGA 240
    DB 307020 ATTGATATGATCTAACTTCCGATGATCATATTATAGATTAAGAATAAGAGACTGAGAGA 307079
    QY 241 ATTTATCAAGATATATGCTTAATCTTTCACTGACATATTAGTGAATGTGATCAA 300

```

DB 307080 ATTATCAAGATATATGCTATATCTTTCATCTGCACATTAATGAGTGAATATGATCAA 307139  
QY 301 ATTCTAAGATATGCAAAATGTTTATTAAGAAATGAATCTGATTTTGAAGGTCAATAT 360  
DB 307140 ATTCTAAGATATGCAAAATGTTTATTAAGAAATGAATCTGATTTTGAAGGTCAATAT 307199  
QY 361 CCTCAAAACATTTGGAATGTTTCCCTGAGCTTGAATTAACCTATGAGCTTATTCAGAT 420  
DB 307200 CCTCAAAACATTTGGAATGTTTCCCTGAGCTTGAATTAACCTATGAGCTTATTCAGAT 307259  
QY 421 GACGATTAATTTATGCACTATATTTTCTCTGTACAGGAATTCACATGAGGAAAT 480  
DB 307260 GACGATTAATTTATGCACTATATTTTCTCTGTACAGGAATTCACATGAGGAAAT 307319  
QY 481 CAACATCAATATGCGCAAGATTTTAAATTAATTAATGATTTCTTATTAATCTTATCTGCT 540  
DB 307320 CAACATCAATATGCGCAAGATTTTAAATTAATTAATGATTTCTTATTAATCTTATCTGCT 307379  
QY 541 GTAATCTTACCTGGGAAAGAGATTTTTCAAAAAATTTTACAAATGATTAAGAGCTTAA 600  
DB 307380 GTAATCTTATGGAAGAGAGATTTTTCAAAAAATTTTACAAATGATTAAGAGCTTAA 307439  
QY 601 TCATTAGAGATTTATTTAGAGAAAAAACTTTCTTAACTTTTCTTGAACGCGCAG 660  
DB 307440 TCATTAGAGATTTATTTAGAGAAAAAATTTCTTAACTTTTCTTGAACGCGCAG 307499  
QY 661 AGATTACCTGATGCGAAGATGATTTTGTGCTGACCAACAGAACGCGCTTAATGAGAG 720  
DB 307500 AGATTACCTGATGCGAAGATGATTTTGTGCTGACCAACAGAACGCGCTTAATGAGAG 307559  
QY 721 GTGACTTTTAAAGACTTAAATTAACAAATCTAGAAATGAGATTTTCTTAATTTGAAAGG 780  
DB 307560 GTGACTTTTAAAGACTTAAATTAACAAATCTAGAAATGAGATTTTCTTAATTTGAAAGG 307619  
QY 781 GCTGCAAAACAAAGATATGATTCATTTATTAAGAGTCAAAAGGTTAACTCTCAACAG 840  
DB 307620 GCTGCAAAACAAAGATATGATTCATTTATTAAGAGTCAAAAGGTTAACTCTCAACAG 307679  
QY 841 ACAGCAGGAAAGATTTGTAACAGCAGTGCAGTAACCTGAAAAATTTGCGAATAT 900  
DB 307680 ACAGCAGGAAAGATTTGTAACAGCAGTGCAGTAACCTGAAAAATTTGCGAATAT 307739  
QY 901 TTAATTAAGTGAAGCTTAAGCCAAAAAGCAGGTTAATCTTAAATGATATCTGAC 960  
DB 307740 TTAATTAAGTGAAGCTTAAGCCAAAAAGCAGGTTAATCTTAAATGATATCTGAC 307799  
QY 961 AATACATGACGGTTCATGATGTTGGAATCTCATTTAATAATATATATGA 1008  
DB 307800 AATACATGACGGTTCATGATGTTGGAATCTCATTTAATAATATATATGA 307847

RESULT 6  
AX770906/c 349980 bp DNA linear PAT 02-JUL-2003  
LOCUS Definition Sequence 37 from Patent WO02094867.  
ACCESSION AX770906  
VERSION AX770906.1 GI:32438070  
KEYWORDS  
SOURCE Photorhabdus luminescens  
ORGANISM Photorhabdus luminescens  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Photorhabdus.  
REFERENCE  
AUTHORS 1  
TITLE Duchaud, E., Taouric, S., Glaser, P., Frangeul, L., Kunst, F.,  
Danchin, A. and Buchrieser, C.  
JOURNAL Patent: WO 02094867-A 37 28-NOV-2002;  
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE  
SCIENTIFIQUE (CNRS) (FR)  
FEATURES  
SOURCE 1. 349980  
/organism="Photorhabdus luminescens"

ORIGIN  
Query Match 98.4%; Score 992; DB 6; Length 349980;  
Best Local Similarity 99.0%; Pred. No. 8e-169;  
Matches 998; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
/mol\_type="genomic DNA"  
/db\_xref="taxon:29488"  
/note="sequence length too big . Split into 2  
sequences.-seq 00037: 1 to 349980-seq 10786: 300001 to  
530312"

QY 1 ATGGTTATACAAATTAACCTGATGATGAGTGAATGATCCACCGCTGAAAGCAATTA 60  
DB 166220 ATGGTTATACAAATTAACCTGATGATGAGTGAATGATCCACCGCTGAAAGCAATTA 166161  
QY 61 GCAGAGATATATAGTATAGTATTAATTAAGCAACAGATGAGGTCATACAGATCA 120  
DB 166160 GCAGAGATATATAGTATAGTATTAATTAAGCAACAGATGAGGTCATACAGATCA 166101  
QY 121 TATGGAATTTGAATTCAGAGCTTAAGAAATTAATTTAGCTTACGCTTGGCTGTAAAGTGT 180  
DB 166100 TATGGAATTTGAATTCAGAGCTTAAGAAATTAATTTAGCTTACGCTTGGCTGTAAAGTGT 166041  
QY 181 ATTCATATGATATCTTAACTTCTGATGATCTATTAATTAAGATTAAGAGAGAGCTGAGAGA 240  
DB 166040 ATTCATATGATATCTTAACTTCTGATGATCTATTAATTAAGATTAAGAGAGAGCTGAGAGA 165981  
QY 241 ATTTATCAAGATATATGATCTTAACTTCTGATGATCTATTAATTAAGATTAAGAGAGAGCTGAGAGA 300  
DB 165980 ATTTATCAAGATATATGATCTTAACTTCTGATGATCTATTAATTAAGATTAAGAGAGAGCTGAGAGA 165921  
QY 301 ATTCTAAGATATGAGCAATGCTTTTATTAAGATGAACCTGATTTTGAAGGTCAATAT 360  
DB 165920 ATTCTAAGATATGAGCAATGCTTTTATTAAGATGAACCTGATTTTGAAGGTCAATAT 165861  
QY 361 CCTCAAAACATTTGGAATGTTCTGAGCTTGAATTAACCAATGAGTGAATTTATTCAGAT 420  
DB 165860 CCTCAAAACATTTGGAATGTTCTGAGCTTGAATTAACCAATGAGTGAATTTATTCAGAT 165801  
QY 421 GACGATTAATTTATGACATATATTTTCTCTGTACAGAAATTCACATGAGGAAAT 480  
DB 165800 GACGATTAATTTATGACATATATTTTCTCTGTACAGAAATTCACATGAGGAAAT 165741  
QY 481 CAACATCAATATGCGCAAGATTTTAAATTAATTAATGATTTCTTATTTAATCTTATCTGCT 540  
DB 165740 CAACATCAATATGCGCAAGATTTTAAATTAATTAATGATTTCTTATTTAATCTTATCTGCT 165681  
QY 541 GTAATCTTACGAGGAGAGATTTTCAAAAACTTTTCAATGATTTAGAGCTTAA 600  
DB 165680 GTAATCTTATGGAAGAGAGATTTTCAAAAACTTTTCAATGATTTAGAGCTTAA 165621  
QY 601 TCATTAGAGATTTATTTAGAGAAAAAACTTTCTTAACTTTCTTTCGACCAACGCGAG 660  
DB 165620 TCATTAGAGATTTATTTAGAGAAAAAAGTTTCTTAACTTTCTTTCGACCAACGCGAG 165561  
QY 661 AGATTACCTGATGCGAAGATGTTTATTTGCTGAGCAACAGAACGCGCTTAATGAGAG 720  
DB 165560 AGATTACCTGATGCGAAGATGTTTATTTGCTGAGCAACAGAACGCGCTTAATGAGAG 165501  
QY 721 GTGAGTTTAAAGACTTAAATTAACAAATCTAGAAATGAGATTTCTTAATTTGAAAGG 780  
DB 165500 GTGAGTTTAAAGACTTAAATTAACAAATCTAGAAATGAGATTTCTTAATTTGAAAGG 165441  
QY 781 GCTGCAAAACAAAGATATGATTCATTTTAAAGAGTACAAAGGTTAAGCTTCAACAG 840  
DB 165440 GCTGCAAAACAAAGATATGATTCATTTTAAAGAGTACAAAGGTTAAGCTTCAACAG 165381  
QY 841 ACAGCAGGAAAGATTTGTAACAGCAGTGCAGTAACTTGAAGAAATTTGCGAATAT 900  
DB 165380 ACAGCAGGAAAGATTTGTAACAGCAGTGCAGTAACTTGAAGAAATTTGCGAATAT 165321  
QY 901 TTAATTAAGTGAAGCTTAAGCCAAAAAGCAGGTTAATCTTAATATGATATCTGAC 960  
DB 165321 TTAATTAAGTGAAGCTTAAGCCAAAAAGCAGGTTAATCTTAATATGATATCTGAC 960

Db 165320 TTATATAGTGTGAGGCTAAGCCAAAAGACAGGGTACCTTCACTCAAAATGATACAG 165261  
Qy 961 AATCAATGACGGTTCATAGTTGGAACTCATTTATAAATATATGA 1008  
Db 165260 AATCAATGACGGTTCATAGTTGGAACTCATTTATAAATATATGA 165213

RESULT 7  
AR407887  
LOCUS AR407887 1107 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 1 from patent US 6630619.  
ACCESSION AR407887  
VERSION AR407887.1 GI:40157855  
KEYWORDS  
SOURCE .  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS East, P.D.  
TITLE Toxin genes from the bacteria *Xenorhabdus nematophilus* and  
JOURNAL *Photobhabdus luminescens*  
FEATURES Patent: US 6630619-A 1 07-OCT-2003;  
source Location/Qualifiers  
1..1107  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 65.9%; Score 663.8; DB 6; Length 1107;  
Best Local Similarity 79.4%; Pred. No. 2.6e-109;  
Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;

Qy 12 ATTAAACCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 71  
Db 57 AGTAAACGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 116  
Qy 72 AGTAAACGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 131  
Db 117 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 176  
Qy 132 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 191  
Db 177 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 236  
Qy 192 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 251  
Db 237 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 296  
Qy 252 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 311  
Db 297 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 356  
Qy 312 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 371  
Db 357 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 416  
Qy 372 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 431  
Db 417 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 476  
Qy 432 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 491  
Db 477 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 536  
Qy 492 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 551  
Db 537 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 596  
Qy 552 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 611  
Db 597 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 656  
Qy 612 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 671

Db 657 TTATATGAGAGAAAAAATTTCTTAACTTTCTTGCACCGCAGAAATGATCTGA 716  
Qy 672 TGGCAGAAATAGATTTTGGCTGACCAAGAGCGCTTAA----- 714  
Db 717 TGGCAGAAATAGATTTTGGCTGACCAAGAGCGCTTAA----- 714  
Qy 715 -----TGGCAGAAATAGATTTTGGCTGACCAAGAGCGCTTAA----- 714  
Db 777 TACAGCAACACGCTACAGCAGCTTCAATTTGAGAGATTTTGCAGAAATCTTAC 836  
Qy 741 AATAGCAACACGCTACAGCAGCTTCAATTTGAGAGATTTTGCAGAAATCTTAC 800  
Db 837 AATAGCAACACGCTACAGCAGCTTCAATTTGAGAGATTTTGCAGAAATCTTAC 896  
Qy 801 TACAGCAACACGCTACAGCAGCTTCAATTTGAGAGATTTTGCAGAAATCTTAC 860  
Db 897 TACAGCAACACGCTACAGCAGCTTCAATTTGAGAGATTTTGCAGAAATCTTAC 956  
Qy 861 TACAGCAACACGCTACAGCAGCTTCAATTTGAGAGATTTTGCAGAAATCTTAC 917  
Db 957 TACAGCAACACGCTACAGCAGCTTCAATTTGAGAGATTTTGCAGAAATCTTAC 1016  
Qy 918 AAGCCAAAAGACAGGGTAACTTCAATTAATGATGATGATGATGATGATGATGAT 977  
Db 1017 AAGCCAAAAGACAGGGTAACTTCAATTAATGATGATGATGATGATGATGATGAT 1076  
Qy 978 TAGGTTGGAATCTTATTAATTAATATATGA 1008  
Db 1077 AAGGTTGGAATCTTATTAATTAATATATGA 1107

RESULT 8  
AX029369  
LOCUS AX029369 1107 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent WO9903328.  
ACCESSION AX029369  
VERSION AX029369.1 GI:10190183  
KEYWORDS  
SOURCE *Xenorhabdus nematophila*  
ORGANISM *Xenorhabdus nematophila*  
REFERENCE 1  
AUTHORS East, P.D.  
TITLE Toxin genes from the bacteria *Xenorhabdus nematophilus* and  
JOURNAL *Photobhabdus luminescens*  
FEATURES Patent: WO 9903328-A 1 28-JAN-1999;  
source COMMUN SCIENT IND RES ORG (AU) ; EAST PETER DAVID (AU)  
1..1107  
/organism="Xenorhabdus nematophila"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:628"

ORIGIN  
Query Match 65.9%; Score 663.8; DB 6; Length 1107;  
Best Local Similarity 79.4%; Pred. No. 2.6e-109;  
Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;

Qy 12 ATTAAACCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 71  
Db 57 AGTAAACGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 116  
Qy 72 AGTAAACGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 131  
Db 117 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 176  
Qy 132 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 191  
Db 177 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 236  
Qy 192 AATAGCTGATGATAGATGATATCCACCCCTTGAAGCAATAGCAGGATAT 251



Db	237	CTCTCACTTCACAGACGCTATTTAAAAATPAGATTAACAAGGTGAATTTATCAAGA	296
Oy	252	ATATATGTCTATCTTTTCATCTGCACATTTAGGTGAATGTGATCAAAATTTCTAAAGA	311
Db	297	ATATACGTCTATCTTTATCTGCACATTTGGGTGAGATGTGTATCAAAATTTCTAAAGA	356
Oy	312	TATGGCAATAGGTTTTTATTAAGATGAACGTGATTTGAAGGTCAATATCTCAAAACAT	371
Db	357	TATGGCAATAGTATTTTACCAGAAACGAACGTGAGTTTGAAGGTCAACGTCTTAAAAATAC	416
Oy	372	TTGGAATGTCCTGAGCTTTGAAAATTAACCATTTGAGTCTTATTCAGATGACGTTAAATT	431
Db	417	CTGGGATATTCCTGATCTTGAGAAATAAACATTTGAAATTTATTCAGATGAAGATTAATT	476
Oy	432	ATTAGCATATATTTTTTCTCTGTACAGAAATTTCACTGGAGGAAAATCAACATCAAA	491
Db	477	ATTAGCATATATTTCTTTGCTTACAGAAACCTTCCATGAGGCAATTCACAAATCAAA	536
Oy	492	TGCGGCAAGATTTTTTAAATTAATTTGATTTCTTATTAACCTTATCTGCTGTAACCTGACT	551
Db	537	TGCAGCAATTTTTTTTAAAGTAATGATTTTACTTATCTTATCTGCTGTAACATGACT	596
Oy	552	GGGAAGAGGATTTTTTCAAAAACTTTTACATGAGATTAGAGCTAATCATTTAGAGA	611
Db	597	GGGAAAAAGGATTTTTTCAAAAAATTTTTCATAGTCTTAAGAACTAATCATTTAGAGA	656
Oy	612	TTATATTAGAGAAAAAAACCTTCTTAAACCTTTCTTTCGACACACGCGAGATTAACCTGA	671
Db	657	TTATATTAGAGAAAAAAACCTTCTTAAACCTTTCTTTCGACACACGCGAGATTAACCTGA	716
Oy	672	TGGCAGAAATGATTATTTTGGCTGACCAACGAACGCGCTAAA-----	714
Db	717	TGGCAGAAACAGGCTACTTGGCGGTCCAACAAAGCGCCTAAATTGCCAACAGCTCTTC	776
Oy	715	-----TGGAGATGAGTTTAAAGACTTAA	740
Db	777	TACAGCAACACGTCCTACAGACCTTCATCTAATTTGAGAGTTAGTTTGGCAAAAACCTTAG	836
Oy	741	AAATTAACAATCTTAGAATGATTTTCTAATATGGAAGGGCTGCAAAAACAAAGTATAG	800
Db	837	AGATTAACCATTCAGAAATTAACATTTATGAATAAGATATGCTGCAAAACGAAATATAG	896
Oy	801	TTCAATTTTAAAGAAGGTACAAAGGGTAAACGCTCCACAGACAGCGAAGAAATATTTGG	860
Db	897	TTCAATTTTAAAGAAGGTACAAAGGGTAAATGATCAAGTCCAGAGACGAAAGTATTTGG	956
Oy	861	TACAGCCAGTGGCAGTAACTGTGAAAAAATTTGCCGATAT--ATTATATTAAGTGTGAGCT	917
Db	957	TACAAAAAAGCGCAGTAACTTCGAAAAACCTCAAGGTGAGATTTATTTAATTAAGACT	1016
Oy	918	AAGCCAAAAGACAGGGGTAACTTTACTCCAAAATGATCTGACAAATCAAGAGAGGTTCA	977
Db	1017	AAGCCAAAAGACAGGGGTAACTTTCTCATTAATTAATTAATCTGACCAAAATTAAGAGATCCA	1076
Oy	978	TAGTGTGGAACCTCATTAATAAATAATATATGA	1008
Db	1077	AAGTCTTGAACCTCATTAACCAAAATATATATA	1107
RESULT 9			
BD080161			
LOCUS			
DEFINITION	BD080161	1107 bp	DNA
			linear
			PAT 27-AUG-2002
ACCESSION			
	BD080161		
VERSION			
	BD080161.1	GI:22625764	
KEYWORDS			
	JP 2001510022-A/1.		
SOURCE			
ORGANISM			
	Xenorhabdus nematophila		
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
	Enterobacteriaceae; Xenorhabdus.		
REFERENCE			
	1 (bases 1 to 1107)		

AUTHORS	East, P. D.
TITLE	Toxin genes from bacteria <i>Xenorhabdus nematophilus</i> and <i>Photorhabdus luminescens</i>
JOURNAL	Patent: JP 2001510022-A 1 31-JUL-2001; COMMUNALHEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
COMMENT	OS <i>Xenorhabdus nematophilus</i> PN JP 2001510022-A/1
PD	31-JUL-2001
PF	17-JUL-1998 JP 2000502652
PR	17-JUL-1997 AU PO 8088
PI	PETER DAVID EAST
PC	C12N1/21, C12N5/10, A01H5/00, A01N63/00, C07K14/24, C12N1/10, C12N1/19 PC
CC	Toxin genes from bacteria <i>Xenorhabdus nematophilus</i> and <i>Photorhabdus luminescens</i>
CC	Photorhabdus
CC	luminescens
FH	Key
FT	source
FT	Location/Qualifiers
FT	1. .1107
FT	/organism="Xenorhabdus nematophilus".
FT	Location/Qualifiers
FT	1. .1107
FT	/organism="Xenorhabdus nematophila"
FT	/mol_type="genomic DNA"
FT	/db_xref="taxon:628"
FEATURES	source
ORIGIN	
Query Match	65.9%; Score 663.8; DB 6; Length 1107;
Best Local Similarity	79.4%; Pred. No. 2.6e-109;
Matches	835; Conservative 0; Mismatches 162; Indels 54; Gaps 2
QY	12 ATTAAACCTGATGATAGATGATATCCACCCGTTGAAAAACCAATAGCAGAGATAT 71
DB	57 AGTAAAGCCTGATGATTAAGGTGAATATCAACCCGTTGAAAAACCAATAGCAGAGATAT 116
QY	72 AGTACGTATCTTAACTTTAAGCAACGATGAGGCTATACAGATCATATGGAATTGA 131
DB	117 AATAGGTGTACTAGAAATTCAGCAACAAAGAAAGTATACAGGATTTGTATGAAATTGC 176
QY	132 ATATGAGCTAAGAAATATATATTTAGCTTACGCTTTGGCTGTAAAGTGTATTCATTAATGT 191
DB	177 ATATGAGCTAAGAAATATATATTTAGCTTACGCTTTGGCTGTAAAGTGTATTCATTAATGT 236
QY	192 ATCTAACTTCTGTATGACTATATTAAGATTAAGAGACTGCTGAGAAATTTATCAAGA 251
DB	237 CTCTCAACTTCCAGAAAGCTATATTAAGATTAAGAGACTGCTGAGAAATTTATCAAGA 296
QY	252 ATATATGCTTAATCTTTGATGCGATTTTAAAGTGAATAATGCTGATCAATTTCTTAAGA 311
DB	297 ATACATGCTTAATCTTTGATGCGATTTTAAAGTGAATAATGCTGATCAATTTCTTAAGA 356
QY	312 TATGGCAATGGTTTTTATTAAGATTAAGAGCTGATTTTAAAGTCAATTCCTCAAAACAT 371
DB	357 TATGGCAATGGTTTTTATTAAGATTAAGAGCTGATTTTAAAGTCAATTCCTCAAAACAT 416
QY	372 TTGAATGTTCTGTAGCTTGAATAATTAACATTTGATGCTTATTCAGATGACGATTAAT 431
DB	417 CTGGATATTCCTGATCTTGAATAATTAACATTTGATGCTTATTCAGATGACGATTAAT 476
QY	432 ATTAGCACTATATTTTTCTCTGACAGAAATTCGACTGAGAGAAAATCAACAATCAAA 491
DB	477 ATTAGCACTATATTTTTCTCTGACAGAAATTCGACTGAGAGAAAATCAACAATCAAA 536
QY	492 TGCCGCAAGTTTTTAAATTAATTAATTTCTATTAATCTTATCTTATCTGTAACCTTC 551
DB	537 TGCCGCAAGTTTTTAAATTAATTAATTTCTATTAATCTTATCTTATCTGTAACCTTC 596
QY	552 GGAAGAGAGATTTTTTCAAAAACTTTTCAATAGATTAAGAGCTAATCAATTAAGAGA 611
DB	597 GGAAGAGAGATTTTTTCAAAAACTTTTCAATAGATTAAGAGCTAATCAATTAAGAGA 656
QY	612 TTATATGAGAGAAAAAATTTCTTAACTTTTCTTGGACCAACGAGAGATTACCTGA 671



Db	657	TTATATTGAGGAGAAAAAACTTTCTTAAACCTTTCTTTGGACACACCGGAGAAAGTTACTGTA	716
Qy	672	TGGCAGAAATVAGSTTATTTGGCTGTGACCAACAGAGCGCTTAA-----	714
Db	717	TGGCAGAACAGGCTACTTGGCCGGTCCAAACAAAGCGCTAAATTGGCAACACACTTTC	776
Qy	715	-----TGAGAGTGAAGTTTAAAGACTTAA	740
Db	777	TACAGCAACAAAGTCTACAGAGACTTCATCTAATTGAGAGAGTTAAGTTTGCACAAAACCTTG	836
Qy	741	AAATTAACAATTTAGGAATGATTTTCTAATATGGAAGGGGTGCACAAACAAAGATATAG	800
Db	837	AGATTAACCATCTCCAAAAATACATTTATGAAAATGATATGCTGCAAAAACGAAAATATAG	896
Qy	801	TTTCATTTTAAAGAGGTAAACAAGGGTAAAGCTCACAGACAGAGGAAAGATATGG	860
Db	897	TTTCATTTTAAAGAGGTAAACAAGGGTAAAGTATCCAGTGCAGAGACAGCAAGATATGG	956
Qy	861	TACAGCAGTGGCAGTAACTCGGAAAAAATTCGCGAATA---ATTATATATAGTGAAGGCT	917
Db	957	TACAAAAACCGGCACTTAACCTTGAAAACTGAGAGGTGAGATTTATATATAGTAAAGCT	1018
Qy	918	AAGCCAAAAAGCAGAGGTAACTTACTCAAAATGATATCTGACATATACATGACGGTCA	977
Db	1017	AAGCCAAAGACAGAGGTAACTTCTCCATATATATATACTGACCAATATATGAGATCCA	1078
Qy	978	TAGTGTGGAATCTCATTTATAAAAATATATGA	1008
Db	1077	AAAGTGTGAAGTCAATTACCAAAATATATATA	1107

LOCUS	AR407889	1205 bp	DNA	linear	PAT 18-DEC-2003
DEFINITION	Sequence 5 from patent US 6630619.				
ACCESSION	AR407889				
VERSION	AR407889.1	GI:40157857			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 1205)				
TITLE	East, P.D.				
JOURNAL	Toxin genes from the bacteria <i>Xenorhabdus nematophilus</i> and				
FEATURES	<i>photorhabdus luminescens</i>				
source	Patent: US 6630619-A 5 07-OCT-2003;				
	Location/Qualifiers				
	1..1205				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	65.9%; Score 663.8; DB 6;	Length 1205;			
Best Local Similarity	79.4%; Pred. No. 2.5e-109;				
Matches	835; Conservative 0; Mismatches 162;	Indels 54; Gaps 2			
12	ATTAAACCTGATGATAGAAAGTGATATCCACCCGTTGAAAACAAATAGCAGAGATAT	71			
11					
10					
9					
8					
73	AGTAACGCTGATGATTAAGGTGAATATCAACCCGTTGAAAAGCAAAATAGCGGAGATAT	132			
72	AGTAGATATACATTAACCTTAAAGCAAAAGATGAGGCTCATAGCATCATATGAAATGA	131			
71					
70					
69					
68					
67					
133	AAATAGCTGATCTAGATTTCAAGCAACCAATGAAATCTCATACGATTTGATAGAAATTC	192			
132					
131					
130					
129					
128					
127					
126					
125					
124					
123					
122					
121					
120					
119					
118					
117					
116					
115					
114					
113					
112					
111					
110					
109					
108					
107					
106					
105					
104					
103					
102					
101					
100					
99					
98					
97					
96					
95					
94					
93					
92					
91					
90					
89					
88					
87					
86					
85					
84					
83					
82					
81					
80					

Db	313	ATACATGCTAATCTTTTATCTGCACATATGGGAGAAATGCTATCAAAATTTCTAAAGA	372
Qy	312	TATGGCAATGCTTTTATATAAGATGAATGGATTTTGAAGTCAATATCTCAAAACAT	371
Db	373	TATGGCAAAATGATTTTACCAGAAACGAATCGAATTTGGAGGTCAAGTCTTAAAAATAC	432
Qy	372	TTGGAAATGTCCTAGCTGGAATAACCATAGAGCTTAATCAGATGACGATTAATY	431
Db	433	CTGGAAATATCTTATCTTGGAATTAACATATGGAATATATTCAGATGAAATTAATY	492
Qy	432	ATTAGCACTATATTTTCTCTGACAGAAATTCACCTGGAGGAAAAATCAACATCAAA	491
Db	493	ATTAGCACTATATTTCTTGCTTACAGAACTTCCATGAGGCAAAATCAACATCAAA	552
Qy	492	TGCGCAAGATTTTAAATTAATGATTTCTTATTTTACCTTATCTGCTTAATCTCACT	551
Db	553	TGCAGCAAAATTTTAAAGTAATGATTTTCTTATCTTATCTGCTTAATCTCACT	612
Qy	552	GGGAAGGAGATTTTTCACAAAACTTTACAAATGATTAAGGCTAAATCATTAAGAA	611
Db	613	GGGAAAAAGATTTTTCACAAAAATTTTACAAATGCTTAACAACTAATCATTAAGAA	672
Qy	612	TTATATGAGAGAAAAAACTTTCTTAAACCTTTCTTTCGACACACGAGATTAACCTGA	671
Db	673	TTATATGAGAGAAAAAACTTTCTTAAACCTTTCTTTCGACACACGAGATTAACCTGA	732
Qy	672	TGGCAGATAGGTTATTTGCTGACCAACAGAAAGCGCTTAA-----	714
Db	733	TGGAGAAACAGGCTACTTGCGCGGTCCAAAGAAAGCGCTTAAATGCAACAGCTTTC	792
Qy	715	-----TGAGAGTGAGTTTAAAGAACTTAA	740
Db	793	TACAGCAACACGCTCTACAGAGCTTCACTTAATTTGGAGTTGAGTTTGGCAAAAACTTAG	852
Qy	741	AAATPACAAATCTAGGAATGGAATTTCTTAATATGGAAGGGGCTGCAAAACAAAGTATAG	800
Db	853	AGATPACCAATCCAGAAATACATTATGAAATGAGATGATCTGCAAAACAAATATAG	912
Qy	801	TTCAATTTTAAAGAGGTACAAAGGGGTAAAGCTCCACAGACAGCGAAAGATATTGG	860
Db	913	TTCAATTTTAAAGAGGTACAAAGGGGTAAAGATTCACAGTCCAGACAGCAAGATTTGG	972
Qy	861	TACAGCCAGTGGCAGTAACTCTGAAAAAATTCGCAAT--ATTATATTAAGTGTGAGCT	917
Db	973	TACAAAAAGCGCAGTAACTTCGAAAAAATCTCAAGTATGAGATTTATTTAATAAGACT	1032
Qy	918	AAAGCCAAAAAGACAGGGTAACTTTATCTCAAAATGATCTGACATACAAATGACGTTCA	977
Db	1033	AAAGCCAAAAAGACAGGGTAACTTTCTCATTAATTAATTACTGACCAAAATATGAGATCCA	1092
Qy	978	TAGTGTGGAATCAATTAATAAATTAATGA	1008
Db	1093	AAAGTGTGGAATCAATTAATAAATTAATGA	1123

RESULT 11					
AX029373					
LOCUS	AX029373	1205 bp	DNA	linear	PAT 16-SEP-2000
DEFINITION	Sequence 5 from Patent WO903328.				
ACCESSION	AX029373				
VERSION	AX029373.1				
KEYWORDS	GI:10190185				
SOURCE					
ORGANISM	Xenorhabdus nematophila				
	Xenorhabdus nematophila				
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
	Enterobacteriaceae; Xenorhabdus.				
REFERENCE	1				
AUTHORS	East, P. D.				
TITLE	Toxin genes from the bacteria xenorhabdus nematophilus and				
	photorhabdus luminescens				
	Patent: NO 9903328-A 5 28-JAN-1999;				
JOURNAL	COMMUN SCIENT IND RES ORG (AU) ; EAST PETER DAVID (AU)				



Db 678 TGGGAAAAAGATTTTTTCAAAAAATTTTCAATGCTCTAGAAACTAAATCATTAGAGA 737  
Oy 611 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 670  
Db 738 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 797  
Oy 671 ATGGCAGAAATAGTTATTTGGCTGGACCAAGAACGCCCTAAA----- 714  
Db 798 ATGGAGAAACAGGCTACTTGGCCGGTCCAAACAAAGCCCTAAATTCGCAACAGCTT 857  
Oy 715 -----TGAGAGTGAGTTT-TAAGAACTT 738  
Db 858 CTACAGCAACAACGCTCTACAGACGCTTCATTAATTGAGAGATTAGTTGCAAAACCTT 917  
Oy 739 AAAAATATCAATCTAGGAATGATTTTCTAATATGAAAGGCGCTGCAAAACAAAGTAT 798  
Db 918 AGAGTTAACCAATCCATCAGAAATACATTTTGAATAATGATGCTGCAAAACAAATAT 977  
Oy 799 AGTTCAATTTATTAAGAGATCAAAAGGATACGCTCCACAGACAGCAACCAAAAGTAT 858  
Db 978 AGTTCAATTTATTAAGAGATCAAAAGGATACGCTCCACAGACAGCAACAGTAT 1037  
Oy 859 GGTACAGCCAGTGGCAGTAACTTGAAAAATTCGCAATA--ATTATATATGTGAGG 915  
Db 1038 GGTACAAAAAGCGGAGTAACTTCGAAAACTGCAAGGTAGAGATTATATATATAGA 1097  
Oy 916 CTAGCCCAAAAAAGAGGTAACCTTTACTCAAAATGATCTGACATACATACGCTT 975  
Db 1098 CTAGCCCAAAAAAGAGGTAACCTTTACTCAAAATGATCTGACATACATATGAGATC 1157  
Oy 976 CATAGTGTGGAACTCATTATTAATAATATATGA 1008  
Db 1158 CAAAGTGTGGAACTCATTACCAAAATATATGA 1190

RESULT 13  
AR082008 834 bp DNA linear PAT 31-AUG-2000  
LOCUS AR082008  
DEFINITION Sequence 3 from patent US 5972687.  
ACCESSION AR082008  
VERSION AR082008.1 GI:10008734  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 834)  
AUTHORS Smigileski, A. Joseph, and Akhurst, R. Joseph.  
TITLE Toxin gene from *Xenorhabdus nematophilus*  
JOURNAL Patent: US 5972687-A 3 26-OCT-1999;  
FEATURES location/Qualifiers  
1..834  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 51.7%; Score 521.4; DB 6; Length 834;  
Best Local Similarity 85.4%; Pred. No. 9.4e-84;  
Matches 604; Conservative 0; Mismatches 101; Indels 2; Gaps 2;

Oy 12 ATTAACACCTGATAGATGATATCCACCCGTTGAAAAAGCAATAGCAGAGATAT 71  
Db 57 AGTAACGCTGATGATTAAGTGAATATCAACCCGTTGAAAAAGCAATAGCGGAGATAT 116  
Oy 72 AGTACGTACTAACTTAAGCAACAGATGAGGCTCATACGATCATATGCAATTGA 131  
Db 117 AATACGTACTAGATTAAGCAACAAATGAAGTCAATCAGATGTTGTAATGCC 176  
Oy 132 ATATGAGCTAAGAAATATATTTAGCTTAGGCTTGTAAGTGAATTCATTAATGT 191  
Db 177 ATATGAGCTAAGAAATATATTTAGCTTAGGCTTGTAAGTGAATTCATTAATGT 236  
Oy 192 ATCTAACTTCTGATGATATATATTAAGAAATGAAGACTGCTGAGAAATTTACAAGA 251

Db 237 CTCTCAACTTCAGAGACATATTAATAAATAGAGATPAACAAGGTAGAAATTTACAGT 296  
Oy 252 ATATATGCTAATCTTTTCATCTGACATATTAGTGAATAATGATCAAAATTTCTAAAGA 311  
Db 297 ATACATGCTAATCTTTTCATCTGACATATTAGTGAATAATGATCAAAATTTCTAAAGA 356  
Oy 312 TATGCAAAATGCTTTTATTAAGAAATGAAGCTGATTTGAAGTCAATATCTCAAAACAT 371  
Db 357 TATGCAAAATGATTTTATCCAGAAACGATGAGCTTTG-AGTCAACGCTTAAATAATAC 415  
Oy 372 TTGAATGTTCTGAGCTTGAAAAATTAACCATT-GAGTGCATTATTCAGATCAATTAAT 430  
Db 416 CTGGAAATATTCCTGATCTTGAATTAACATTAATGAAGATTTATTCAGATCAATTAAT 475  
Oy 431 TATTAAGCATATATTTTCTGTAAGAAATTTCCACTGAGAAATCAACATCA 490  
Db 476 TATTAAGCATATATTTCTGTAAGAAATTTCCACTGAGAAATCAACATCA 535  
Oy 491 ATGCGGAGATTTTAAATTAATTAATTAATTTCTTAATTAATCTGCTGTAACCTTCA 550  
Db 536 ATGAGCAAAATTTTAAATTAATTAATTAATTTCTTAATTAATCTGCTGTAACCTTCA 595  
Oy 551 TGGAGAGAGATTTTCAAAAACTTTTCAATGATTAAGGCTTAATCATTTAGAGA 610  
Db 596 TGGAAAAAGGATTTTCAAAAAATTTTACAATGCTTAAGAACTTAATCATTTAGAGA 655  
Oy 611 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 670  
Db 656 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 715  
Oy 671 ATGGCAGAAATAGTTATTTGGCTGGACCAAGAGAGGCCCTAAATGG 717  
Db 716 ATGGCAGAAACAGGCTACTTGGCCGGTCCAAACAAAGCCCTAAATGG 762

RESULT 14  
AR082007 837 bp DNA linear PAT 31-AUG-2000  
LOCUS AR082007  
DEFINITION Sequence 1 from patent US 5972687.  
ACCESSION AR082007  
VERSION AR082007.1 GI:10008733  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 837)  
AUTHORS Smigileski, A. Joseph, and Akhurst, R. Joseph.  
TITLE Toxin gene from *Xenorhabdus nematophilus*  
JOURNAL Patent: US 5972687-A 1 26-OCT-1999;  
FEATURES location/Qualifiers  
1..837  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 51.7%; Score 521.4; DB 6; Length 837;  
Best Local Similarity 85.4%; Pred. No. 9.4e-84;  
Matches 604; Conservative 0; Mismatches 101; Indels 2; Gaps 2;

Oy 12 ATTAACACCTGATAGATGATATCCACCCGTTGAAAAAGCAATAGCAGAGATAT 71  
Db 57 AGTAACGCTGATGATTAAGTGAATATCAACCCGTTGAAAAAGCAATAGCGGAGATAT 116  
Oy 72 AGTACGTACTAACTTAAGCAACAGATGAGGCTCATACGATCATATGCAATTGA 131  
Db 117 AATACGTACTAGATTAAGCAACAAATGAAGTCAATCAGATGTTGTAATGCC 176  
Oy 132 ATATGAGCTAAGAAATATATTTAGCTTAGGCTTGTAAGTGAATTCATTAATGT 191  
Db 177 ATATGAGCTAAGAAATATATTTAGCTTAGGCTTGTAAGTGAATTCATTAATGT 236  
Oy 192 ATCTAACTTCTGATGATATATATTAAGAAATGAAGACTGCTGAGAAATTTACAAGA 251

Db 237 CTCGACCTTCAGAAAGACTATTATATAAAGATACACAGGATGATTTATCAAGT 296  
Qy 252 ATATATGCTAATCTTTCATCTGCACATTAAGTGAAATGATGATCAAAATTTCTAAAGA 311  
Db 297 ATACATGCTAATCTTTCATCTGCACATTAAGTGAAATGATGATCAAAATTTCTAAAGA 356  
Qy 312 TATGCAAAATGCTTTCATTAAGATGAATGATTTGAAAGGTCATATCCCAAAACAT 371  
Db 357 TATGCAAAATGATTTTACCCAGACGAACTGAGTTG-AGGTCAACGCTTAAAAATAC 415  
Qy 372 TTGATGATGCTGAGCTTGAATTAACCATTT-GAGTCTTATTCAGATGAGATTAAT 430  
Db 416 CTGGATATCTCTGATCTTGAGATTAATTAATGAAAGTATTTATTCAGATGAGATTAAT 475  
Qy 431 TATTAACCATTAATTTTCTCTGTACAGAAATTCACCTGAGAGAAATCAACATCA 490  
Db 476 TATTAACCATTAATTTTCTCTGTACAGAAATTCACCTGAGAGAAATCAACATCA 535  
Qy 491 ATGCGGCAAGATTTTTCATTAATTAATGATTTCTTATTAACCTTATCTGCTTAACCTTC 550  
Db 536 ATGAGCAAAATTTTTCATTAATTAATGATTTTCTTATTAACCTTATCTGCTTAACCTTC 595  
Qy 551 TGGGAGAGAGATTTTTCATTAATTAATGATTTTCTTATTAACCTTATCTGCTTAACCTTC 610  
Db 596 TGGGAGAGAGATTTTTCATTAATTAATGATTTTCTTATTAACCTTATCTGCTTAACCTTC 655  
Qy 611 ATTATATTTGAG 670  
Db 656 ATTATATTTGAG 715  
Qy 671 ATGGAG 717  
Db 716 ATGGAG 762

RESULT 15  
AX083744/c 1141 bp DNA linear PAT 28-FEB-2001  
LOCUS AX083744 Sequence 22 from Patent W00111061.  
DEFINITION AX083744  
ACCESSION AX083744  
VERSION AX083744.1 GI:13185472  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
synthetic construct  
artificial sequences.  
REFERENCE  
1 Kunst, L. and Clemens, S.  
TITLE Regulation of embryonic transcription in plants  
JOURNAL Patent: WO 011061-A 22 15-FEB-2001;  
UNIVERSITY OF BRITISH COLUMBIA (CA)  
FEATURES  
source  
1. 1141  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
1. 1141  
/note="consensus sequence of A.t., L.a., and B.n. FAE1  
promoters"  
ORIGIN  
Query Match 6.2%; Score 62.6; DB 6; Length 1141;  
Best Local Similarity 11.1%; Pred. No. 0.16;  
Matches 81; Conservative 280; Mismatches 360; Indels 6; Gaps 1;  
Qy 185 ATATGATCTAATCTTCTGATGACTATTATAAGATTAAGAGACTGCTGAGAGATTT 244  
Db 834 ACRNTRTWABWKSWMNN 775  
Qy 245 ATCAAGATATATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 304  
Db 774 NNYYMAATVTTHTDCTKTMWMTWYMDWMTTMBTTTNNMTTSTNNNNNNNNNNNNNNNN 715  
Qy 305 CTAAAGATATGCAAAATGGTTTATTAAGATGAACCTGATTTTGAAGTCAATATCTCTC 364

Db 714 NNNNNKAYAHATNNWGCMMNNNTDARTNNNTYVRRRWMTNTKRWSTYRRHHYTAT 655  
Qy 365 AAAACATTTGAAATGCTTCGAGCTTGAATAAATCAATGAGTGCATTT-----CAG 418  
Db 654 NNN 595  
Qy 419 ATGACGATTAATTTATTAACATTAATTTTCTCTGTACAGAAATTTCACTGAGAGAA 478  
Db 594 WMYANWMCBDVYTRNNNTYCKSYAHSYWYNNNAWYRRYSABWSSMAATWTRNNMM 535  
Qy 479 ATCAACATTCAAATGCCCAAGATTTTTCATTAATTAATGATTTCTTATTAACCTTATCTG 538  
Db 534 SGBVWRVAGTMMWMMHNNNNNTDTRYYWWRWABRTTYVDSMCNAXSMWRGNWRMK 475  
Qy 539 CTGTAATCTTCATGAGAGAGAGATTTTTCATTAATTAATGATTTTCAATGATTAAGAGCTA 598  
Db 474 MWAANNDAGAMDHTYTMGNNTYMMRRAMWMMAMWBRAYCCNNNNNNRACVWHKHGM 415  
Qy 599 AATCATTAAGAGATTAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658  
Db 414 RWTWKYMKKAACNNNNBRAMWYRVAMWVYSDTTNTIDMMWMTSDMBMHWYTVDYTMRAW 355  
Qy 659 AGAGATTAACCTGATGCGAGATTAAGTTATTTGGCTGAGACCAACAGAGCGCTAAATGA 718  
Db 354 NNNNNNNNRBCKTYSMMWMDHNNTHCTYGNNWTGSAVBMWAMSMWAGAASBVYNNCW 295  
Qy 719 GAGTGAATTTTAAAGAACTTAATAAATCAATCTAGAGATGATTTTCTAATATGAAG 778  
Db 294 RMTYMGKMTNN 235  
Qy 779 GGGCTGCAAAACAAAGATATGTTCAATTAATTAAGAGAGTCAAAAGGTTAACGCTCAC 838  
Db 234 GKHHWMBWRABHRSNNNNWVYKCRNKYVSWHYHAMRYBKVABAIVGCNNNNKDBAHHHW 175  
Qy 839 AGACAGCAGCAAAAGATTTTGTACAGCATGTGACAGTACCTGAGAAATTTGCCGATA 898  
Db 174 CATNNNNMMWYAYVHHNHKKKAAWNTNTABRDBDHAHVKTYYWYWDYWCAMCWMA 115  
Qy 899 ATTATTA 905  
Db 114 KAKYRTA 108

Search completed: November 21, 2004, 07:40:05  
Job time : 4346 secs